

USO DE MÉTODOS QUANTITATIVOS PARA IDENTIFICAÇÃO DE *CLUSTERS*
DE PALAVRAS EM UM *THESAURUS* COM APLICAÇÕES EM TRAÇOS DE
PERSONALIDADE

Alfredo Ricardo de Faria Passos

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Produção.

Orientadores: Marcos Pereira Estellita Lins

Luiz Biondi Neto

Luiz Amâncio Machado de Souza
Junior

Rio de Janeiro
Novembro de 2014

USO DE MÉTODOS QUANTITATIVOS PARA IDENTIFICAÇÃO DE *CLUSTERS*
DE PALAVRAS EM UM *THESAURUS* COM APLICAÇÕES EM TRAÇOS DE
PERSONALIDADE

Alfredo Ricardo de Faria Passos

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM
CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO.

Examinada por:

Prof. Marcos Pereira Estellita Lins, D. Sc.

Prof. Luiz Biondi Neto, D. Sc.

Prof. Luiz Amâncio Machado de Souza Junior, D. Sc.

Prof. Luiz Pereira Calôba, D. Sc.

Prof. Reinaldo Morabito Neto, D. Sc.

Prof. Marcelo Gerardin Poirot Land, D. Sc.

Prof. Edilson Fernandes de Arruda, D. Sc.

RIO DE JANEIRO, RJ – BRASIL

NOVEMBRO DE 2014

Passos, Alfredo Ricardo de Faria

Uso de métodos quantitativos para identificação de clusters de palavras em um thesaurus com aplicações em traços de personalidade / Alfredo Ricardo de Faria Passos. - Rio de Janeiro: UFRJ/COPPE, 2014.

XVII, 332 p.: il.; 29,7 cm.

Orientadores: Marcos Pereira Estellita Lins

Luiz Biondi Neto

Luiz Amâncio Machado de Souza Junior

Tese (Doutorado) – UFRJ/ COPPE/ Programa de Engenharia de Produção, 2014.

Referências Bibliográficas: p. 283-299.

1. Clusters. 2. Traços de Personalidade. 3. Avaliação Psicológica I. Lins, Marcos Pereira Estellita *et al* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Produção. III. Título.

Ao meu pai, Edilmar Passos,
Ao meu irmão, Vitor,
à minha mãe, Icléa,

AGRADECIMENTOS

Ao meu pai, em pensamento, meu herói em vida e símbolo de sabedoria e honestidade.

Ao meu irmão Vitor, pelo exemplo de flexibilidade, e felicidade com a vida.

A minha mãe, Icléa, por me mostrar o caminho da vida e de se estar de bem com a vida.

Ao professor Marcos Estellita pelos conhecimentos transmitidos, esclarecimentos, pelas preciosas orientações no desenvolvimento deste trabalho, por ter acreditado em mim no momento em que aceitou me orientar e por me mostrar a importância da metacognição no esforço de nos tornarmos melhores seres humanos.

Ao professor Luiz Biondi pelo preciso suporte na área de redes neurais de Kohonen, pela receptividade e preciosos conselhos.

Ao professor Luiz Pereira Calôba, pelas suas preciosas orientações, que ajudaram fundamentalmente em alguns dos desenvolvimentos deste trabalho.

Ao professor Luiz Amâncio pela orientação e disponibilização de estruturas e recursos para o meu trabalho.

A minha segunda mãe Cremilda, pelo símbolo de garra, caráter e dedicação às pessoas que ama.

Ao professor Luis Otávio Façanha, amigo e debatedor de temas ligados à econometria no IE UFRJ.

Aos meus saudosos professores do Colégio de Aplicação, pela contribuição acadêmica e me ajudarem na formação como cidadão, em especial Gustavo Bernardo, Cristina Ferreira e João Carlos Cataldo.

Ao amigo Márcio, pelo apoio nos momentos difíceis e pelas conversas francas.

Aos amigos Walter Júnior, Leandro Marino, Felipe Galúcio, Samir Thompson e Daniela Devos.

Ao professor Olavo de Carvalho por me apresentar um mundo inteiro novo de conhecimento.

A Deus, por me conceder inteligência e energia para realizar minha missão no universo.

Resumo da Dissertação apresentada à COPPE / UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D. Sc.)

USO DE MÉTODOS QUANTITATIVOS PARA IDENTIFICAÇÃO DE *CLUSTERS*
DE PALAVRAS EM UM *THESAURUS* COM APLICAÇÕES EM TRAÇOS DE
PERSONALIDADE

Alfredo Ricardo de Faria Passos

Novembro/2014

Orientadores: Marcos Pereira Estellita Lins

Luiz Biondi Neto

Luiz Amâncio Machado de Souza Junior

Programa: Engenharia de Produção

A avaliação psicológica de indivíduos se constitui um tema de extrema relevância para as diversas organizações humanas e para as práticas clínicas. Os modelos que buscam explicar as diferenças individuais tiveram, na linguagem natural, a principal fonte de oferta de termos descritores dos diferentes comportamentos humanos.

Por outro lado, os desenvolvimentos desses modelos de personalidade se deram por meio de aplicações de técnicas de clusterização (ou seja, técnicas de obtenção de agrupamentos) de dados e de análise fatorial; através das quais se conseguiu identificar um pequeno número de variáveis/fatores capazes de explicar boa parte dessas diferenças individuais. Esta tese desenvolveu uma série de rotinas (e um *software* específico) capazes de, a partir do *True Term Thesaurus Database*, gerar tais relatórios e programar algoritmos de clusterização em conjuntos de traços de personalidade.

Nestes esforços desenvolveu um algoritmo inovador de codificação dos vértices de um grafo como vetores; de forma que estes últimos possam servir de *input* a um dos algoritmos de clusterização utilizados (redes neurais SOM).

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

USE OF QUANTITATIVE METHODS IN THE IDENTIFICATION OF WORD
CLUSTERS IN A THESAURUS WITH APPLICATIONS IN PERSONALITY
TRAITS

Alfredo Ricardo de Faria Passos

November/2014

Advisors: Marcos Estellita Pereira Lins
Luiz Biondi Neto
Luiz Amâncio Machado de Souza Junior

Department: Program of Production Engineering

The *Psychological* evaluations of individuals constitute an extremely relevant theme for human organizations as a whole and for the clinical practices. The models that seek to explain those individual differences had, on the natural language, its main source of terms descriptors of the different human behaviors.

On the other hand, the development of these personality models took place with the application of clusterization techniques or factor analysis; by means of which it was possible to identify a specific set of traits/factors capable to explain a great degree of these individual differences. This thesis developed useful reports and provided trait clusterization solutions to the professionals and researchers in the personality field have importance to these professionals.

In these efforts, one developed an innovative graph node codification algorithm that codifies the nodes of a graph as vectors, so that these structures can serve as input for one of the utilized clusterization algorithms: SOM neural networks.

SUMÁRIO

AGRADECIMENTOS	V
SUMÁRIO	VIII
LISTA DE FIGURAS	XIII
LISTA DE TABELAS	ERRO! INDICADOR NÃO DEFINIDO.
1 INTRODUÇÃO	1
1.1 CONTEXTUALIZAÇÃO DA SITUAÇÃO-PROBLEMA, RELEVÂNCIA E JUSTIFICATIVO DO ESTUDO	1
1.2 MOTIVAÇÕES DO ESTUDO.....	4
1.3 CONTRIBUIÇÕES E DESENVOLVIMENTOS RELEVANTES	6
1.3.1 CONTRIBUIÇÕES ORIGINAIS.....	6
1.3.2 DESENVOLVIMENTOS RELEVANTES.....	8
1.4 OBJETIVOS DO ESTUDO.....	9
1.5 PERGUNTAS A SEREM RESPONDIDAS NESTA TESE.....	10
1.5.1 PODEM ALGORITMOS EM GRAFOS, EM PARTICULAR DE CLUSTERIZAÇÃO, AUXILIAREM NA PESQUISA E PRÁTICA RELACIONADAS À PERSONALIDADE HUMANA?.....	10
1.5.2É POSSÍVEL OFERECER MÉTODOS QUE POSSIBILITEM A UTILIZAÇÃO DE ALGORITMOS DE CLUSTERIZAÇÃO CONHECIDOS COMO REDES NEURAIS DE KOHONEN PARA SEREM APLICADOS NOS PROBLEMAS DE CLUSTERIZAÇÃO EM GRAFOS?	10
1.5.3 PODEM AS TÉCNICAS AQUI DESENVOLVIDAS FORMAR UMA BASE PARA O DESENVOLVIMENTO DE UMA FERRAMENTA GERENCIAL QUE AUXILIE NAS DIVERSAS FUNÇÕES ORGANIZACIONAIS ENVOLVENDO PERSONALITY ASSESSMENT?.....	10
1.6 ORGANIZAÇÃO DO DOCUMENTO	10
2 MÉTODOS QUANTATIVOS NA DETERMINAÇÃO DE CLUSTERS	12
2.1 O PROBLEMA GERAL DE CLUSTERIZAÇÃO.....	12
2.1.1 .. CLUSTERIZAÇÃO HIERÁRQUICA.....	14
2.1.1.1 MEDIDAS DE DISTÂNCIA UTILIZADAS POR DIFERENTES ALGORITMOS DE CLUSTERIZAÇÃO.....	15
2.1.1.2 CRITÉRIOS DE LIGAÇÃO COMUMENTE ADOTADOS POR ALGORITMOS DE CLUSTERIZAÇÃO HIERÁRQUICA	16
2.1.1.3 OUTROS ALGORITMOS AGLOMERATIVOS/DIVISIVOS DE CLUSTERIZAÇÃO.....	17
2.1.2 CLUSTERIZAÇÃO BASEADA EM CENTRÓIDE (ALGORITMO <i>K-MEANS</i>	17
2.1.3 CLUSTERIZAÇÃO BASEADA EM DENSIDADES.....	18
2.1.4 MÉTODOS DE CLUSTERIZAÇÃO DE DADOS BASEADOS EM ANÁLISE FATORIAL.....	18
2.1.5 OUTROS MÉTODOS DE CLUSTERIZAÇÃO.....	18

2.2 ALGORITMOS DE CLUSTERIZAÇÃO DE GRAFOS	19
2.2.1 APRENDIZADO NÃO-SUPERVISIONADO EM PROBLEMAS DE CLUSTERIZAÇÃO.....	21
2.2.1.1 REDES NEURAS ARTIFICIAIS DE KOHONEN (OU SOM) NADAYARA-WATSON.....	21
2.2.1.2 REDES SOM EM CLUSTERIZAÇÃO DE ÁRVORES E FLORESTAS.....	26
2.2.1.3 METODOLOGIA PROPOSTA PARA CODIFICAÇÃO DOS VÉRTICES DE UM GRAFO (ÁRVORE OU FLORESTA) VISANDO-SE APLICAÇÕES DE CLUSTERIZAÇÃO.....	27
2.2.1.4 OUTRAS CARACTERÍSTICAS DA METODOLOGIA DE CODIFICAÇÃO PROPOSTA.....	42
2.2.1.5 A CODIFICAÇÃO PROPOSTA E A MALDIÇÃO DA DIMENSIONALIDADE.....	43
2.2.1.6 A MALDIÇÃO DA DIMENSIONALIDADE.....	43
2.2.1.7 OUTROS PROBLEMAS IDENTIFICADOS COM A MALDIÇÃO DA DIMENSIONALIDADE...	44
2.2.1.8 CONSEQUÊNCIAS DA MALDIÇÃO DA DIMENSIONALIDADE PARA ALGORITMOS QUE UTILIZEM A CODIFICAÇÃO PROPOSTA DOS VÉRTICES DE UM GRAFO ACÍCLICO.....	46
2.2.1.9 APLICAÇÃO DA METODOLOGIA PROPOSTA EM GRAFOS COM CICLOS.....	47
2.2.2.0 EXEMPLO DE APLICAÇÃO DO ALGORITMO DE CODIFICAÇÃO PROPOSTO EM UM GRAFO CONTENDO CICLOS QUE COMPARTILHAM VÉRTICES.....	47
2.2.2.1 UM EXEMPLO DE CÓDIGO EM JAVA APLICADO AO ALGORITMO DE CODIFICAÇÃO....	51
2.2.2.2 CONSIDERAÇÕES FINAIS SOBRE O ALGORITMO PROPOSTO.....	51
2.2.2.3 ESQUEMAS ALTERNATIVOS DE UTILIZAÇÃO DO ALGORITMO PROPOSTO EM PROBLEMAS DE CLUSTERIZAÇÃO DE GRAFOS.....	52
2.2.3 ALGORITMO DE MARKOV <i>CLUSTERING</i> (MCL).....	52
2.2.3.1 EXEMPLO ILUSTRATIVO DO FUNCIONAMENTO DO ALGORITMO MCL.....	55
2.2.4 OUTROS ALGORITMOS DE CLUSTERIZAÇÃO EM GRAFOS.....	58
2.2.4.1 <i>K-MEANS</i> DE VETORES DE ALTA DIMENSÃO.....	58
2.2.4.2 HEURÍSTICAS E META-HEURÍSTICAS NA CLUSTERIZAÇÃO EM GRAFOS.....	58
2.2.4.3 SPECTRAL CLUSTERING.....	59
2.2.4.4 CLUSTERIZAÇÃO UTILIZANDO ELEMENTOS DE TEORIA DE GRAFOS.....	59
2.2.4.5 CLUSTERIZAÇÃO BASEADA EM CLIQUES EM GRAFOS.....	60
2.2.4.6 ALGORITMO DE FLUXO MÁXIMO/CORTE MÍNIMO.....	60
2.3 ALGUMAS APLICAÇÕES ANTERIORES DE CLUSTERIZAÇÃO ENVOLVENDO TRAÇOS DE PERSONALIDADE.....	60
2.3.1 OUTROS ESTUDOS DE CLUSTERIZAÇÃO ENVOLVENDO TRAÇOS DE PERSONALIDADE A CITAR.....	61
3 CONSIDERAÇÕES SOBRE MÉTODOS DE CLUSTERIZAÇÃO DE PALAVRAS EM THESAURI.....	62

3.1 NOTAS IMPORTANTES SOBRE A ESTRUTURA DE RELAÇÕES ENTRE PALAVRAS EM UM THESAURUS.....	67
3.1.1 EXEMPLOS DE QUE NEM TODAS AS ASSOCIAÇÕES CUMPREM FUNÇÕES DE SINÔNIMOS OU ANTÔNIMOS.....	68
4 ALGUNS DOS PRINCIPAIS DESENVOLVIMENTOS NA ÁREA DE TRAÇOS DE PERSONALIDADE ANTERIORMENTE FEITOS.....	70
4.1 CONCEITOS E CONCEPÇÕES ACERCA DE TRAÇOS DE PERSONALIDADE....	70
4.1.1 CONCEPÇÕES CIENTÍFICAS DE TP.....	71
4.2 REVISÃO DOS DESENVOLVIMENTOS EM TRAÇOS DE PERSONALIDADE....	73
4.2.1 CONDIÇÕES NECESSÁRIAS PARA O DESENVOLVIMENTO DA PESQUISA CIENTÍFICA DE TRAÇOS.....	74
4.2.2 OS PRIMEIROS ESTUDOS EMPÍRICOS.....	74
4.2.2.1 .. ESTUDO DE ALLPORT E ODBERT (1936).....	75
4.2.2.2 ESTUDOS DE RAYMOND CATELL.....	77
4.2.2.2.1 A IDENTIFICAÇÃO DE 171 TRAÇOS DE PERSONALIDADE COM BASE EM ANÁLISE FATORIAL.....	77
4.2.2.2.2 O DESENVOLVIMENTO DO QUESTIONÁRIO DOS 16 FATORES DA PERSONALIDADE.....	85
4.2.2.2.3 ESTUDO DE NORMAN (1963)	89
4.2.2.2.4 ESTUDO DE TUPES E CRISTAL (1961).....	91
4.2.2.3 OS MODELOS DE CINCO FATORES DE PERSONALIDADE.....	93
4.2.2.4 EQUIVALÊNCIAS ENTRE OS DIFERENTES MODELOS DE CINCO FATORES DA PERSONALIDADE.....	95
4.2.2.5 INFLUÊNCIA DE FATORES GENÉTICOS	96
4.2.2.6 . INFLUÊNCIA DAS DIFERENÇAS CULTURAIS SOBRE OS CINCO FATORES.....	96
4.2.2.7 CRÍTICAS AOS MODELOS DE CINCO FATORES DA PERSONALIDADE.....	97
5 SISTEMAS DESENVOLVIDOS, BANCOS DE DADOS, ROTINAS DESENVOLVIDAS, PROGRAMAS UTILIZADOS E CARACTERÍSTICAS DO THESAURUS ADOTADO.....	ERRO! INDICADOR NÃO DEFINIDO.
5.1 ESTRUTURA DO BANCO EM PLANILHA EXCEL	99
5.2 LINGUAGENS DE PROGRAMAÇÃO UTILIZADAS.....	101
5.3 BANCOS DE DADOS UTILIZADOS	101
5.4 OUTROS SOFTWARES UTILIZADOS.....	101
5.5 SISTEMA DESENVOLVIDO EM LINGUAGEM PHP.....	102
5.5.1 SISTEMA PHP	104
5.5.2 INTERFACES GRÁFICAS DO SISTEMA EM PHP	106

5.5.3	EXEMPLOS DE RELATÓRIOS GERADOS PELO SISTEMA EM PHP	113
5.5.4	MELHORAMENTOS POSSÍVEIS DE SEREM FEITOS NO SISTEMA EM PHP.....	113
5.5.5	ROTINAS EM JAVA	115
5.5.5.1	ROTINA DE IMPORTAÇÃO DE DADOS DO TTT EM PLANILHA EXCEL.....	116
5.5.5.2	ROTINA DE GERAÇÃO DE LISTAS DE ADJACÊNCIA A PARTIR DE DADOS IMPORTADOS.....	116
5.5.5.2.1	PROGRAMAÇÃO ORIENTADA A OBJETO NO BANCO DE DADOS <i>MySQL</i>	117
5.5.5.3	ROTINAS DE GERAÇÃO DE MATRIZES.....	127
5.5.5.4	ROTINAS RELACIONADAS AO ALGORITMO MCL E RATINGS DE CLUSTERS ENCONTRADOS PELO MCL.....	127
5.5.5.4.1	ROTINAS RELATIVAS ÀS FUNCIONALIDADES DO ALGORITMO MCL (MARKOV <i>CLUSTERING</i>).....	127
5.5.6	ROTINAS GERADORAS DE LISTAS DE ADJACÊNCIA.....	128
5.6	<i>HARDWARES</i> UTILIZADOS.....	128
5.7	RELAÇÃO ENTRE RELATÓRIOS GERADOS, ALGORITMOS DE CLUSTERIZAÇÃO E CTPT'S	128
6	DETALHES DE IMPLEMENTAÇÃO COM DIFERENTES ALGORITMOS DE CLUSTERIZAÇÃO.....	131
6.1	CARACTERÍSTICAS DO CONJUNTO DE PALAVRAS SELECIONADAS	132
6.1.1	CRITÉRIOS PARA SUBSTITUIÇÃO DE ALGUMAS DAS 171 PALAVRAS DE CATTELL, POR PALAVRAS SIMILARES NO <i>THESAURUS</i> UTILIZADO	146
6.1.2	ANÁLISE DO GRAFO QUE REPRESENTA AS ASSOCIAÇÕES ENTRE AS 222 PALAVRAS SELECIONADAS.....	152
6.2	DETALHES DE IMPLEMENTAÇÃO.....	163
6.2.1	REDES SOM.....	165
6.2.2	MCL (MARKOV <i>CLUSTERING</i>).....	168
7	RESULTADOS DAS CLUSTERIZAÇÕES NO CONJUNTO DE 222 TRAÇOS.....	170
7.1	RESULTADOS COM REDES SOM	170
7.2	RESULTADOS COM MCL (MARKOV CLUSTERING).....	190
7.3	COMPARAÇÃO DOS RESULTADOS OBTIDOS ENTRE AS DIFERENTES METODOLOGIAS	199
7.4	COMPARAÇÃO DOS RESULTADOS DE CLUSTERIZAÇÃO OBTIDOS COM O MODELO DE 12 FATORES DE PERSONALIDADE DE RAYMOND CATTELL (1945B).....	204
7.5	NOVA CONFIGURAÇÃO E TREINAMENTO DA REDE SOM	240
7.5.1	RESULTADOS ENCONTRADOS.....	241

8	APLICABILIDADES DAS ROTINAS COMPUTACIONAIS IMPLEMENTADAS EM ÁREAS RELACIONADAS À TRAÇOS DE PERSONALIDADE.....	250
8.1	BREVE REVISÃO DAS MOTIVAÇÕES PARA O DESENVOLVIMENTO DAS ROTINAS COMPUTACIONAIS E DO SISTEMA PHP.....	250
8.2	BREVE REVISÃO DOS RELATÓRIOS GERADOS PELAS ROTINAS E PELO SISTEMA DESENVOLVIDOS.....	252
8.2.1	OBSERVAÇÕES GERAIS SOBRE OS DESENVOLVIMENTOS.....	253
8.2.1.1	CONSIDERAÇÕES GERAIS SOBRE AS ATRATIVIDADES DAS SOLUÇÕES COMPUTACIONAIS IMPLEMENTADAS	255
8.3	APLICABILIDADES DAS ROTINAS E SISTEMA DESENVOLVIDOS.....	256
8.3.1	APLICAÇÕES PSICOLÓGICAS	256
8.3.1.1	APLICAÇÕES PSICOLÓGICAS/ORGANIZACIONAIS	256
8.3.1.1.1	RECRUTAMENTO E SELEÇÃO.....	257
8.3.1.1.1.1	EXEMPLO DE UTILIZAÇÃO DAS ROTINAS DESENVOLVIDAS EM UMA APLICAÇÃO ORGANIZACIONAL (FICTÍCIA) ENVOLVENDO UM PROCESSO SELETIVO.....	260
8.3.1.1.1.2	SEGUNDO EXEMPLO DE UTILIZAÇÃO DAS ROTINAS DESENVOLVIDAS EM UMA APLICAÇÃO ORGANIZACIONAL (FICTÍCIA) ENVOLVENDO UM PROCESSO SELETIVO	264
8.3.1.1.2	PROCEDIMENTOS DE AVALIAÇÃO SISTEMÁTICA E MELHORIAS INDIVIDUAIS DOS FUNCIONÁRIOS.....	269
8.3.1.1.3	FORMAÇÃO DE GRUPOS DE TRABALHO AFINS.....	269
8.3.1.2	FORMAÇÃO DE GRUPOS DE TRABALHO AFINS.....	269
8.3.1.3	APLICAÇÕES NA ÁREA DE PESQUISA EM TEORIA DE TRAÇOS DE PERSONALIDADE.....	272
8.3.2	OUTRAS APLICAÇÕES	273
8.3.2.1	APLICAÇÕES LINGUÍSTICAS E ENSINO DE LÍNGUAS	273
8.3.2.2	APLICAÇÕES NA ÁREA DE <i>KNOWLEDGE MANAGEMENT</i>	274
8.3.2.3	OUTRAS APLICAÇÕES EM TEORIA DE GRAFOS/ALGORÍTMOS EM GRAFOS	274
9	CONCLUSÕES E RECOMENDAÇÕES DE ESTUDOS FUTUROS.....	276
9.1	CONCLUSÕES.....	276
9.2	RECOMENDAÇÕES DE ESTUDOS FUTUROS.....	281
	REFERÊNCIAS BIBLIOGRÁFICAS	283
	APÊNDICE.....	300
	ANEXO	332

LISTA DE FIGURAS

Figura 1: Esquemas gráficos de clusterização de vetores e em grafos ..	19
Figura 2: Exemplo de topologia hexagonal de um mapa auto-organizável ..	22
Figura 3: Exemplo de topologia aleatória de um mapa auto-organizável ..	22
Figura 4: Exemplo de identificação inicial de cada dado de entrada com os modelos que melhor os representam ..	23
Figura 5: Esquema ilustrativo das etapas 1), 2) e 3) do algoritmo de codificação ..	30
Figura 6: Grafo com dois componentes conexos, utilizado de exemplo do algoritmo de codificação dos vértices de um grafo ..	32
Figura 7: Identificação da 1-vizinhança no 1º componente conexo do grafo ..	33
Figura 8: LVPC após inclusão dos vértices 2, 3, e 4 ..	34
Figura 9: Identificação dos vértices adjacentes ao 1-vizinho 2 no 1º componente conexo do grafo ..	34
Figura 10: LVPC após inclusão do vértice 5 ..	35
Figura 11: LVPC após inclusão dos vértices 6 e 8 ..	35
Figura 12: Identificação da 8-vizinhança (vértices vizinhos ao vértice 8) no 1º componente conexo do grafo ..	36
Figura 13: LVPC após inclusão dos vértices 10, 20 e 50 ..	36
Figura 14: LVPC após a inclusão de todos os vértices do 1º componente conexo do grafo ..	37
Figura 15: Identificação da 20-vizinhança na 2ª componente conexa ..	38
Figura 16: LVPC após inclusão de todos os vértices da 2ª componente conexa do grafo ..	39
Figura 17: LVPC após final do processo de codificação ..	40
Figura 18: Árvore composta por cinco vértices ..	41
Figura 19: LVPC após o final da codificação do exemplo 2 ..	42
Figura 20: Configuração dos ciclos do grafo a ser utilizado no algoritmo de codificação ..	47
Figura 21: Vetores representando os 16 vértices do grafo utilizado como exemplo, ao final do processo de codificação ..	48
Figura 22: Correspondência entre os 16 fatores identificados por Cattell e as cinco dimensões amplas da personalidade de Cattell ..	89
Figura 23: Esquema de estruturação do <i>thesaurus</i> em inglês ..	99
Figura 24: Tela de Entrada do Sistema PHP ..	107
Figura 25: Tela de Importação de Palavras do Sistema em PHP ..	108
Figura 26: Tela de Listas no Sistema em PHP ..	109
Figura 27: Tela de Criação de Listas no Sistema em PHP ..	110
Figura 28: Tela de Geração de Matriz de Adjacência do Sistema em PHP ..	111
Figura 29: Tela de Geração dos Demais Relatórios do Sistema em PHP ..	112

Figura 30: Tela de <i>Download</i> dos Relatórios Gerados pelo Sistema em PHP	112
Figura 31: Diagrama Conceitual de Classes Utilizadas na Programação a Objeto no Banco de Dados <i>My SQL</i>	117
Figura 32: Alocação final no mapa topológico utilizado na rede SOM	174
Figura 33: Relação entre relatórios gerados, pelo Sistema em PHP e pelas rotinas em Java e os algoritmos de clusterização utilizados com o objetivo de se determinar CTPT's.....	129
Figura 34: Alocação dos traços da 1ª componente conexa do grafo no mapa topológico.....	246
Figura 35: <i>U-Matrix</i> da nova clusterização dos traços na componente conexa 1 do grafo.....	248
Figura 36: <i>Cluster 1</i> – Metodologia SOM.....	300
Figura 37: <i>Cluster 2</i> – Metodologia SOM.....	301
Figura 18: <i>Cluster 3</i> – Metodologia SOM.....	302
Figura 39: <i>Cluster 4</i> – Metodologia SOM.....	302
Figura 40 <i>Cluster 5</i> – Metodologia SOM.....	303
Figura 41: <i>Cluster 6</i> – Metodologia SOM.....	303
Figura 42: <i>Cluster 7</i> – Metodologia SOM.....	304
Figura 43: <i>Cluster 8</i> – Metodologia SOM.....	304
Figura 44: <i>Cluster 14</i> – Metodologia SOM.....	305
Figura 45: <i>Cluster 15</i> – Metodologia SOM.....	305
Figura 46: <i>Cluster 16</i> – Metodologia SOM.....	306
Figura 47: <i>Cluster 17</i> – Metodologia SOM.....	306
Figura 48: <i>Cluster 18</i> – Metodologia SOM.....	307
Figura 49: <i>Cluster 19</i> – Metodologia SOM.....	307
Figura 50: <i>Cluster 20</i> – Metodologia SOM.....	308
Figura 51: <i>Cluster 21</i> – Metodologia SOM.....	308
Figura 52: <i>Cluster 23</i> – Metodologia SOM.....	309
Figura 53: <i>Cluster 28</i> – Metodologia SOM.....	309
Figura 54: <i>Cluster 29</i> – Metodologia SOM.....	310
Figura 55: <i>Cluster 33</i> – Metodologia SOM.....	310
Figura 56: <i>Cluster 34</i> – Metodologia SOM.....	311
Figura 57: <i>Cluster 43</i> – Metodologia SOM.....	311
Figura 58: <i>Cluster 44</i> – Metodologia SOM.....	312
Figura 59: <i>Cluster 45</i> – Metodologia SOM.....	312
Figura 60: <i>Cluster 46</i> – Metodologia SOM.....	313
Figura 61: <i>Cluster 48</i> – Metodologia SOM.....	313
Figura 62: <i>Cluster 59</i> – Metodologia SOM.....	314

Figura 63: <i>Cluster</i> 61 – Metodologia SOM.....	314
Figura 64: <i>Cluster</i> 62 – Metodologia SOM.....	315
Figura 65: <i>Cluster</i> 71 – Metodologia SOM.....	315
Figura 66: <i>Cluster</i> 72 – Metodologia SOM.....	316
Figura 67: <i>Cluster</i> 75 – Metodologia SOM.....	316
Figura 68: <i>Cluster</i> 85 – Metodologia SOM.....	317
Figura 69: <i>Cluster</i> 97 – Metodologia SOM.....	317
Figura 70: <i>Cluster</i> 99 – Metodologia SOM.....	318
Figura 71: <i>Cluster</i> 1 – Metodologia MCL.....	318
Figura 72: <i>Cluster</i> 2 – Metodologia MCL.....	319
Figura 73: <i>Cluster</i> 3 – Metodologia MCL.....	319
Figura 74: <i>Cluster</i> 4 – Metodologia MCL.....	320
Figura 75: <i>Cluster</i> 5 – Metodologia MCL.....	320
Figura 76: <i>Cluster</i> 6 – Metodologia MCL.....	321
Figura 77: <i>Cluster</i> 7 – Metodologia MCL.....	321
Figura 78: <i>Cluster</i> 8 – Metodologia MCL.....	322
Figura 79: <i>Cluster</i> 9 – Metodologia MCL.....	322
Figura 80: <i>Cluster</i> 10 – Metodologia MCL.....	323
Figura 81: <i>Cluster</i> 11 – Metodologia MCL.....	323
Figura 82: <i>Cluster</i> 12 – Metodologia MCL.....	324
Figura 83: <i>Cluster</i> 13 – Metodologia MCL.....	324
Figura 84: <i>Cluster</i> 14 – Metodologia MCL.....	325
Figura 85: <i>Cluster</i> 15 – Metodologia MCL.....	325
Figura 86: <i>Cluster</i> 16 – Metodologia MCL.....	326
Figura 87: <i>Cluster</i> 21 – Metodologia MCL.....	326
Figura 88: <i>Cluster</i> 22 – Metodologia MCL.....	327
Figura 89: <i>Cluster</i> 25 – Metodologia MCL.....	327
Figura 90: <i>Cluster</i> 26 – Metodologia MCL.....	328
Figura 91: <i>Cluster</i> 29 – Metodologia MCL.....	328
Figura 92: <i>Cluster</i> 30 – Metodologia MCL.....	329
Figura 93: <i>Cluster</i> 32 – Metodologia MCL.....	329
Figura 94: <i>Cluster</i> 33 – Metodologia MCL.....	330
Figura 95: <i>Cluster</i> 41 – Metodologia MCL.....	330
Figura 96: <i>Cluster</i> 50 – Metodologia MCL.....	331
Figura 97: <i>Cluster</i> 61 – Metodologia MCL.....	331

LISTA DE TABELAS

Figura 18: <i>Cluster 3</i> – Metodologia SOM.....	302
.....	xiv
Tabela 26: Alocação das palavras nas regiões do mapa topológico.....	174
.....	xvi
Tabela 37: Percentuais de ligações <i>intra</i> e <i>inter clusters</i> verificados para os resultados obtidos com o algoritmo MCL na componente conexa 1 do grafo.....	187 ... xvi
Tabela 16: Os Cinco Fatores de Norman.....	89
Tabela 17: Presença comparada das 171 variáveis selecionadas por Catell em diversos modelos de fatores de personalidade.....	90
Tabela 18: Modelo dos Cinco Grandes Fatores de Tupes e Cristal.....	92
Tabela 19: Modelo dos Cinco Grandes Fatores da Personalidade de acordo com MACRAE e OLIVER (1992).....	94
Tabela 20: Alinhamento entre três dos principais modelos de cinco fatores de personalidade.....	96
Tabela 21: Exemplo de lista de adjacência gerada por interface de geração de lista de adjacência em java com 222 palavras.....	118
Tabela 22: As 222 palavras selecionadas para a aplicação das metodologias de clusterização.....	133
Tabela 23: As 222 palavras consideradas e seus respectivos componentes conexos.....	153
Tabela 24: Número de palavras por componente conexo do grafo.....	160
Tabela 25: Resultados obtidos com redes de Kohonen a partir das 222 <i>Keywords</i>	170
Tabela 26: Alocação das palavras nas regiões do mapa topológico.....	174
Tabela 37: Percentuais de ligações <i>intra</i> e <i>inter clusters</i> verificados para os resultados obtidos com o algoritmo MCL na componente conexa 1 do grafo.....	187
Tabela 28: Percentuais de ligações <i>intra</i> e <i>inter clusters</i> verificados para os resultados obtidos com o algoritmo MCL nas demais componentes conexas do grafo.....	189
Tabela 29: <i>Clusters</i> obtidos com o algoritmo MCL com parâmetro de <i>inflation</i> = 1,85.....	190
Tabela 30: Percentuais de ligações <i>intra</i> e <i>interclusters</i> verificados para os resultados obtidos com o algoritmo MCL na componente conexa 1 do grafo.....	197
Tabela 31: Percentuais de ligações <i>intra</i> e <i>interclusters</i> verificados para os resultados obtidos com o algoritmo MCL nas demais componentes conexas do grafo.....	198
Tabela 32: Comparação entre os clusters obtidos com redes SOM e MCL.....	199

Tabela 33: Comparação dos fatores do modelo 12 PF de Cattell (1945b), com os clusters obtidos com o algoritmo MCL.....	207
Tabela 34: Comparação dos fatores do modelo 12 PF de Cattell (1945b), com os clusters obtidos com o algoritmo SOM.....	223
Tabela 35: Resultados encontrados na nova configuração SOM na componente conexa 1do grafo.....	242
Tabela 36: Percentuais de ligações <i>intra</i> e <i>inter clusters</i> na 1ª componente conexa do grafo.....	246
Tabela 37: Relatório de <i>clusters</i> do algoritmo MCL gerado por rotina em Java.....	261
Tabela 38: Exemplo de questionário gerado a partir do relatório de <i>clusters</i> , por rotina em Java e preenchido para fins de ilustração (contendo somente os traços selecionados e seus antônimos).....	264
Tabela 39: Documento ilustrativo de texto contendo entrevista (transcrita) com participante de processo seletivo.....	265
Tabela 40: Relatório de identificação da frequência de traços em textos de avaliação de um determinado indivíduo.....	268
Tabela 41: Carta do linguista da empresa True Term Thesaurus Database, recebida em 25-11-2013 confirmando que amajoria das ligações diretas entre palavras neste thesaurus correspondem a sinônimos.....	332

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO DA SITUAÇÃO-PROBLEMA, RELEVÂNCIA E JUSTIFICATIVA DO ESTUDO

A importância do emprego de *clusters* de palavras em repositórios do tipo *thesaurus* tem aumentado consideravelmente recentemente, tanto em termos de aplicações práticas, como por exemplo, utilização desses *clusters* no desenvolvimento de mecanismos de busca de conteúdos na internet como na organização de documentos, *knowledge management* etc. Também se tem observado o uso de tais *clusters* no aprendizado de línguas e no aumento do vocabulário em idiomas, dentre outras aplicações.

Como será visto mais adiante, *clusters* (classes ou subconjuntos) são agrupamentos de objetos tais que, objetos reunidos em um mesmo agrupamento possuem maior similaridade entre si (segundo um critério pré-definido) do que entre objetos pertencentes a diferentes agrupamentos.

Em termos de desenvolvimentos acadêmicos, verifica-se um grande número de trabalhos nas áreas de organização do conhecimento por similaridade do conteúdo de textos (como em PARVIN, MINAEI-BIDGOLI e DAHBASHI, 2012), desenvolvimentos de mecanismos de buscas na *internet* (BRASLAVSKI, ALSHANSKI e SHISHKIN, 2004), dentre outros.

Uma das principais razões do crescente uso de tais *clusters*, em diferentes áreas do conhecimento reside no fato de que tais agrupamentos de palavras, definidos em *thesauri* deste tipo, apresentam propriedades de extrema relevância para propósitos de organização das mesmas:

-Em cada *cluster* é possível se identificar a(s) ideia(s) /conceito(s) geral (is) presente(s) nas palavras a ele pertencentes. Esse conceito geral pode servir como uma forma de rotulação desses agrupamentos. Assim, uma clusterização efetuada em tal base de dados propiciará uma organização de palavras, na qual cada agrupamento trará vocábulos organizados e agrupados segundo determinadas ideias mais gerais, que os associa.

- Adicionalmente, as palavras diretamente relacionadas umas as outras, em cada conjunto deste tipo, cumprem na grande maioria dos casos, papel de sinônimas entre si; fato que contribui para que esses conjuntos sejam uma forma útil de organização das mesmas dentro das diversas aplicações que fazem uso destes dispositivos.

Deste modo, em cada *cluster* identificado, para cada par de palavras a ele pertencentes e diretamente relacionado umas com as outras se teria, na esmagadora maior parte dos casos, sinônimos. Assim, na maior parte dos casos verifica-se que associações diretas entre palavras nestes *thesauri*, relacionadas nestes *clusters*, cumprem o papel de sinônimos.

Assim, estas características fornecem uma base para imediata catalogação e organização, ao menos inicial, desta base de dados.

Por outro lado, a teoria de traços de personalidade, por sua vez, tem sido objeto de atenção de diversos trabalhos acadêmicos e de diversas aplicações organizacionais, como por exemplo, em processos de *personality assessment* (a esse respeito vide KRUG, 1989).

O número de trabalhos desenvolvidos neste campo revela a importância que estes traços têm na vida humana, seja na melhor compreensão do indivíduo ou ainda em processos de avaliação de indivíduos, na formação de grupos de trabalho afins, etc. Trabalhos recentes mostram a importância desses traços até mesmo na identificação de líderes e suas contribuições em problemas sociais complexos (CAVICCHIOLI, 2008). Uma apresentação de problemas sociais complexos pode ser encontrada em LINS e DINIZ (2012).

Contudo, especialmente no que se refere às aplicações organizacionais da teoria de traços de personalidade, não se encontrou na literatura a existência de soluções algorítmicas de clusterização de palavras em *thesauri* para fornecimento, aos profissionais de áreas de interesse (como por exemplo, profissionais de departamentos pessoais, recursos humanos, gerentes responsáveis por equipes de trabalho, psicólogos organizacionais etc.) relatórios que possam subsidiar algumas de suas tarefas, como por exemplo, recrutamento e seleção, montagem de grupos de trabalho, avaliação psicológica de indivíduos e de grupos de indivíduos etc. Também se verifica tal ausência nas áreas clínicas (até o momento da conclusão deste trabalho só se obteve conhecimento de um único desenvolvimento, que utilizava um thesaurus físico no fornecimento de soluções de apoio às práticas clínicas; trata-se do thesaurus clínico de ZUCKERMAN, 2010).

Levando-se em consideração que a identificação de características comportamentais relacionadas a aspectos da personalidade presentes nos indivíduos, assim como a elaboração de métodos de avaliação psicológica são questões de grande importância no dia-a-dia das organizações, este trabalho desenvolveu um conjunto de soluções computacionais que visam auxiliar estes profissionais nestas tarefas.

Como as mencionadas características desses *clusters* em *thesauri* deste tipo fornecem uma base para organização de palavras quaisquer, pode-se também encontrar *clusters* de

traços de personalidade quando se utilizam, nestes repositórios, somente palavras que correspondam a traços.

Similarmente, estas mesmas características desses *clusters* podem auxiliar no desenvolvimento de diversas outras aplicações que envolvam os resultados de um procedimento de clusterização de palavras em dicionários deste tipo.

Tais *clusters* podem prover uma grande gama de serviços às profissionais de áreas correlatas incluindo a possibilidade de auxiliar o pesquisador da área na seleção de traços de interesse, em cada *cluster* identificado. Pode assim auxiliar tal profissional na seleção daqueles traços que cubram, ou representem, da forma mais abrangente possível, dependendo das aplicações, os aspectos da personalidade presentes em cada *cluster*.

Tal auxílio é possível pelas próprias características de *clusters* de palavras em *thesaurus* (CPT) uma vez que, nos mesmos, os vocábulos são agrupados em subconjuntos nos quais sempre há sentidos/ideias mais amplas referindo-se a todos eles e que, adicionalmente, palavras diretamente relacionadas são sinônimas entre si na grande maioria dos casos.

Com vistas à determinação de tais agrupamentos de traços fazia-se necessário a implementação de algoritmos de clusterização que pudessem implementar formas sistemáticas na determinação desses conjuntos.

Com isto visa-se também contribuir de forma útil aos profissionais da área que passariam a contar com tais métodos sistemáticos na determinação desses úteis agrupamentos.

Com este intuito foram desenvolvidos e/ou implementados diversos algoritmos de grafos e de clusterização em grafos aplicados a um repositório específico de palavras do tipo *thesaurus*.

Isto se tornou necessário, pois, as determinações desses agrupamentos de interesse envolvem a necessidade de se modelar a rede de relacionamentos entre tais palavras como um grafo; e, portanto, tais agrupamentos, teriam que ser determinados por algoritmos quantitativos de clusterização em tais estruturas.

Deste modo, esta tese aplica métodos quantitativos que visam encontrar agrupamentos de traços, segundo o *framework* de uma abordagem psicoléxica, em um *thesaurus* específico. Com isto, visa suprir este tipo de abordagem na determinação destes agrupamentos com métodos sistemáticos e que propiciem flexibilidade aos potenciais usuários destas ferramentas.

Para tanto, utiliza-se de duas técnicas distintas na determinação desses agrupamentos de palavras: Redes Neurais Artificiais de Kohonen (ou *Self Organizing Maps*, SOM) e o algoritmo conhecido como *Markov Clustering* (MCL), proposto por DONGEN (2000).

1.2 MOTIVAÇÕES DO ESTUDO

O primeiro fato que motivou os desenvolvimentos deste estudo foi à verificação de que diversos estudos psicoléxicos pioneiros na área de traços de personalidade partiram de uma seleção de palavras em um dicionário (físico) Webster para a seguir catalogarem manualmente e exaustivamente os traços em categorias que refletiam a similaridade entre os mesmos no que se refere aos aspectos comportamentais por eles representados.

Assim, os trabalhos pioneiros desenvolvidos na abordagem psicoléxica (especialmente por ALLPORT e ODBERT, 1936 e por CATTELL, 1943b) recorreram a tais procedimentos para, com base em desenvolvimentos posteriores – no caso de Catell e outros estudos – chegarem a um conjunto menor de traços e, a partir desses, procederem à construção de um modelo de grandes fatores da personalidade. Assim, CATTELL (1943a, 1943b, 1956 e 2000), termina por chegar a um modelo de dezesseis grandes fatores da personalidade. Outro desenvolvimento subsequente utilizou-se de conjuntos similares de variáveis, incluindo o modelo do *Big Five Factors of Personality* (BFFP).

No momento atual, alguns trabalhos acadêmicos (como em JACKSON e PAUNONEN, 2000) desenvolvidos nos últimos anos têm apontado que ainda há espaço para pesquisa na área; incluindo a área de modelos de fatores de personalidade.

Nesta área, há evidências empíricas de que o número de fatores - e de variáveis compondo cada fator – varia em diferentes culturas. JACKSON e PAUNONEN (2000) mostram que a própria organização desses fatores varia de cultura em cultura. Mesmo o modelo mais utilizado na área, o modelo dos Cinco Grandes Fatores da Personalidade, ou *Big Five*, BFFP tem sido criticado a esse respeito.

Assim, verificou-se que a pesquisa na área de traços de personalidade apresenta espaço para novas contribuições, que visem a superar algumas das limitações acima apresentadas. Neste sentido, dentro de um *framework* de um esforço de pesquisa na área, baseado na hipótese léxica, que se baseia na premissa de que as diferenças em personalidade podem ser representadas em termos de linguagem natural, torna-se desejável um mecanismo que possibilite organizar os traços de personalidade presentes em um grande banco de dados,

como por exemplo, um dicionário e que esta organização possa auxiliar de alguma forma a pesquisa e práticas profissionais relacionadas a traços.

Um esquema possível de organização destes traços seria, por exemplo, agrupá-los de acordo com as similaridades entre si segundo a linguagem natural.

Finalmente, este método de organização de traços deveria ser de fácil e de relativamente rápido uso, evitando-se que tempo demasiado seja gasto nestas etapas.

Adicionalmente, a identificação de termos relacionados podem ter outras aplicações específicas em Psicologia, como por exemplo, apresentar uma grande gama de termos relacionados visando melhorar a clareza e precisão na confecção e redação de relatórios psicológicos. ZUCKERMAN (2010) apresenta um *thesaurus* (físico) clínico, na forma de livro, contendo termos clínicos específicos para suporte na elaboração de relatórios deste tipo, como por exemplo, na elaboração e relatórios de clientes em terapias psicológicas, contendo seus progressos e demais assuntos relacionados. Assim percebeu-se a oportunidade de se prover termos relacionados (e em especial traços de personalidade) por meio de *clusters* de traços de personalidade em thesauri (CTPT).

Ainda na área clínica, percebeu-se que uma ferramenta que pudesse fornecer ao clínico, uma forma de levantamento, na forma de relatórios, das palavras (e especialmente, dos traços de personalidade) mais presentes nos relatos e nas redações dos pacientes se poderia auxiliar este profissional na identificação de características relevantes desses pacientes.

Assim cogitou-se a possibilidade de desenvolver ferramentas computacionais que, para um determinado conjunto de termos selecionados, apresentassem relatórios com esse enfoque promovendo a contagem de todos esses termos de interesse em um dado relatório, efetuado pelo clínico.

Além disso, seria desejável em práticas profissionais desse tipo, que para cada termo constante em um relatório, se pudessem apresentar relatórios adicionais que explicitassem os seus respectivos *clusters*; podendo-se inclusive, incluir uma grande quantidade de palavras do *thesaurus* utilizado, nestes relatórios.

Desta forma a rápida identificação de *clusters*, com a utilização de soluções de clusterização eficientes, passa a ser de extrema relevância para os pesquisadores e/ou profissionais que desejem realizar experimentos, ou aplicações, baseados em CTPT. Notou-se ainda a necessidade de se utilizar algoritmos capazes de encontrar *clusters* para grandes conjuntos de traços, incluindo a possibilidade de se utilizar de todos os traços presentes em um *thesaurus*, por meio do algoritmo MCL.

Outra contribuição deste trabalho é fornecer uma abordagem sistemática e algorítmica que possa auxiliar também na seleção e organização das palavras em geral contidas em um *thesaurus* específico, em *clusters*.

Essa praticidade auxilia o pesquisador que segue uma abordagem psicoléxica de traços de personalidade de várias formas:

Em primeiro lugar auxilia o pesquisador, ou profissional trabalhando em uma determinada organização, a de forma mais rápida e sistemática, selecionar as palavras (em cada *cluster* encontrado) que representam de forma mais abrangente e exaustiva, os aspectos de personalidade de interesse em uma determinada aplicação e que se encontram representados por cada *cluster* obtido.

A poupança de tempo e esforço na seleção de adjetivos traços, através da leitura de um dicionário por meio das rotinas computacionais, de clusterização e de elaboração de relatórios relacionados aos clusters encontrados representa ganho de produtividade para este público.

Com estas motivações em mente se procedeu aos desenvolvimentos deste trabalho

1.3 CONTRIBUIÇÕES E DESENVOLVIMENTOS RELEVANTES

Este trabalho realizou uma série de contribuições em diversas áreas a serem listados a seguir.

1.3.1 Contribuições originais

- Na área da Pesquisa Operacional, desenvolveu-se de forma inovadora um algoritmo de codificação dos vértices de grafos para que se possa utilizar o algoritmo de Redes neurais de Kohonen ou o de *K-means* na determinação de *clusters* nos mesmos. O algoritmo proposto codifica cada vértice do grafo como um vetor de alta dimensão.

Tal algoritmo, quando utilizado em grafos acíclicos, gera vetores tais que, quando utilizados por um dos dois mencionados algoritmos de clusterização, os quadrados das distâncias euclidianas computadas entre cada par de vértices do grafo, em um mesmo componente conexo, correspondem ao exato número de passos (em termos de arestas do grafo percorridas) necessários para se atingir um desses vértices partindo-se do outro.

Em grafos cíclicos, tal metodologia aproxima a distância entre pares de vértices, pertencentes a um determinado ciclo, determinando as exatas menores distâncias para todos os demais vértices em um determinado componente conexo do grafo.

- Ainda na área de Pesquisa Operacional *hard* (que se ocupa da aplicação de métodos formais quantitativos na modelagem e solução de problemas organizacionais), aplicou de

forma inovadora uma rede neural SOM com a codificação proposta acima em um problema de clusterização de grafos.

- Na área de algoritmos voltados para a determinação de CPT, aplicou-se, de forma inédita, o algoritmo de redes neurais de Kohonen, conjuntamente com a metodologia de codificação de vértices de um grafo proposta.

- Ainda nesta área, aplicou de forma inovadora o algoritmo MCL diretamente no *thesaurus* utilizado dentro de um contexto de traços de personalidade.

-Na área de prática clínica oferece soluções algorítmicas inéditas que visam auxiliar os profissionais envolvidos nestas práticas, através de um conjunto de rotinas, associadas aos *clusters* encontrados pelo algoritmo MCL, por meio da elaboração de diversos relatórios úteis para tal prática. Isto se dá de duas formas:

a) Na elaboração de relatórios que facilitem a elaboração de relatórios clínicos, por meio do fornecimento de termos relacionados em sentido com um determinado termo de interesse. Neste sentido, os *clusters* de palavras em *thesauri* (CPT), quando apresentados na forma de relatórios, podem cumprir tais propósitos. Estas possibilidades estão presentes para quaisquer CPT, inclusive os compostos exclusivamente por traços de personalidade.

Por meio destes expedientes, o clínico pode escolher os termos mais apropriados para descrever o progresso de determinado paciente e, portanto, na melhor confecção destes relatórios.

b) Na elaboração de relatórios que facilitem profissionais e pesquisadores da área de traços na escolha de um conjunto de traços de interesse, uma vez que os *clusters* de traços de personalidade em *thesauri* (CTPT) proveem a organização dos mesmos em conjuntos com as características dos CPT já mencionadas.

Uma vez que uma das formas de se promover o *assessment* da personalidade de determinado indivíduo é através da análise dos seus relatos verbais ou das redações por ele elaboradas, desenvolveram-se de forma inédita, rotinas que promovem a contagem do número de vezes que cada vocábulo aparece nos seus relatos (ou em suas redações ou ainda, presentes em gravações de registros sobre pacientes feitos pelo próprio clínico) associando-se, adicionalmente, a contagem de cada um desses termos com o *cluster* a que pertence.

Tais rotinas corroboram nas tarefas do clínico referentes ao posicionamento psicológico dos seus pacientes e ao progresso dos mesmos em terapias. Estas contagens, como se verá, podem ser obtidas a partir dos meios mencionados, desde que elas sejam transformadas em documentos de texto com extensões específicas.

Também fornecem grande flexibilidade de trabalho a este público, pois permitem ao mesmo checar se determinados traços existem no *thesaurus* em questão e escolher quais traços irão ser utilizados nas rotinas computacionais e nos relatórios relacionados ao grafo intrínseco aos traços selecionados.

Em particular, as rotinas de clusterização aplicadas a esta base de dados, fornecem de modo sistemático e atrativo de organização inicial destes traços, organizados por similaridade, de acordo com a linguagem natural. Estes *clusters*, ao agruparem termos segundo a linguagem natural podem fornecer auxílio importante na identificação de dimensões da personalidade humana, poupando o pesquisador de uma busca arbitrária e/ou manual desses traços em um dicionário comum e da posterior tarefa de organizá-los por similaridade quanto aos aspectos comportamentais a que se referem.

As rotinas aqui desenvolvidas permitem aos profissionais da área utilizar o conjunto de possibilidades que tais rotinas proporcionam desde a seleção de traços de interesse, passando pela determinação de *clusters* desses traços até a elaboração de relatórios baseados nos mesmos.

Assim, com base nos desenvolvimentos mencionados acima, visa proporcionar aos pesquisadores e profissionais da área flexibilidade de trabalho e pesquisa, incluindo alguns ambientes computacionais amigáveis (no sistema desenvolvido em linguagem PHP) de computação e também favorecendo um ambiente de aprendizagem a partir dos dados e dos resultados de clusterização obtidos.

- Na área de *software*, desenvolveu de modo inédito um programa capaz de ler o *True Term Thesaurus Database* (TTT) e gerar alguns relatórios de grafos subjacentes aos relacionamentos entre conjuntos de palavras de interesse: matrizes de adjacência, listas de adjacência e relatório de relação entre palavras e rótulos numéricos; desenvolvido em linguagem PHP). Estes relatórios são muito importantes tanto para uma compreensão inicial deste grafo como servindo de input pra outros algoritmos.

- Finalmente a tese promove, de forma também inédita, uma discussão entre os diferentes critérios que podem ser levados em consideração (ou combinados) para a determinação de *clusters*.

1.3.2 Desenvolvimentos relevantes

Além das contribuições acima elencadas, este trabalho também alcançou outras realizações que, embora não sejam propriamente contribuições originais, são importantes para outras áreas do conhecimento, a saber:

- Utilização pela primeira vez, de aplicações computacionais com o *thesaurus* produzido pela empresa TT-Software/Databases (daqui para frente se utilizará a sigla TTT para se referir a este repositório).

- Implementação de rotinas em linguagem Java relacionadas à elaboração de relatórios relacionados à Teoria de Grafos e aos algoritmos de clusterização, usando o algoritmo MCL, capazes de serem aplicadas a grandes quantidades de palavras no *thesaurus* utilizado. Alguns desses relatórios serão apresentados ao longo do desenvolvimento desta tese e outros estão disponíveis no apêndice do mesmo.

1.4 OBJETIVOS DO ESTUDO

Este trabalho pretende cumprir os seguintes objetivos:

- Contribuir para a facilitação da pesquisa, práticas clínicas e outras aplicações profissionais envolvendo traços de personalidade por meio do provimento de algoritmos e rotinas computacionais que forneçam relatórios úteis nestes campos e, sobretudo esquemas de organização de traços de personalidade segundo a linguagem comum (a partir das relações entre estes traços tal como identificada pelo *True Term Thesaurus Database*).

Esta contribuição se baseia tanto no fornecimento de soluções que propiciem flexibilidade na pesquisa e prática na área, como na disponibilização de um algoritmo eficiente de clusterização em grafos (MCL).

Estas contribuições se baseiam em:

Encontrar, por meio de algoritmos de clusterização, agrupamentos de traços de personalidade, relacionados segundo a linguagem natural e contando com as propriedades dos CPT, já apresentadas anteriormente.

Para tanto e propõe a aplicar, dois algoritmos: redes neurais Kohonen e algoritmo *Markov Clustering*.

b) Construir outros relatórios relacionados ao grafo subjacente a um determinado conjunto de palavras selecionadas neste *theasurus*.

-Na área de grafos e de Pesquisa Operacional *Hard*, se propõe a desenvolver um algoritmo de codificação de vértices de grafos para que possam ser utilizados pelos algoritmos de clusterização de redes neurais SOM e *K-means*. Como será apresentado, tal algoritmo possibilita a aplicação destas máquinas de aprendizado em problemas de clusterização de grafos onde só se disponha de informação relativa às adjacências entre os vértices dos mesmos, de uma forma simples.

Assim, este trabalho visa contribuir com a área de traços de personalidade fornecendo métodos e rotinas flexíveis e integradas, que possam tanto auxiliar pesquisadores como outros profissionais que se utilizem da teoria de traços e necessitem de rotinas integradas que forneçam uma série de relatórios, dentro da abordagem psicológica.

1.5 PERGUNTAS A SEREM RESPONDIDAS NESTA TESE

Esta tese se propõe a responder as seguintes perguntas:

1.5.1 Podem algoritmos em grafos, em particular de clusterização, auxiliarem na pesquisa e prática relacionadas à personalidade humana?

1.5.2 Ainda com relação aos métodos de clusterização em grafos, é possível oferecer métodos que possibilitem a utilização de algoritmos de clusterização conhecidos como redes neurais de Kohonen para serem aplicados nos problemas de clusterização em grafos?

1.5.3 As técnicas aqui desenvolvidas podem formar uma base para o desenvolvimento de uma ferramenta gerencial que auxilie nas diversas funções organizacionais envolvendo *personality assessment*?

1.6 ORGANIZAÇÃO DO DOCUMENTO

Este trabalho está organizado em nove capítulos. O primeiro capítulo é constituído das motivações, objetivos e perguntas a serem respondidas por este estudo. Neste mesmo capítulo se contextualiza e se apresenta a relevância do estudo, assim como as contribuições e desenvolvimentos relevantes para as áreas de clusterização de dados, Pesquisa Operacional e traços de personalidade.

O capítulo dois promove uma breve revisão de diversos tipos de problemas de clusterização e dos algoritmos desenvolvidos para abordá-los. Este capítulo depois foca no problema de clusterização de grafos e apresenta a contribuição original feita na área (com um algoritmo de codificação de vértices de um grafo para que possa ser utilizado por redes neurais de Kohonen e por algoritmos *K-means*), para um problema específico de clusterização de grafos.

O capítulo três trata de importantes considerações acerca de *clusters* de palavras em *thesauri*, de características desses agrupamentos e apresenta particularidades de como o *thesaurus* utilizado está estruturado.

O capítulo quatro traz importantes considerações sobre traços de personalidade, das etapas necessárias para se construir cientificamente modelos de personalidade e faz um histórico dos principais desenvolvimentos na área.

O capítulo cinco apresenta o *software* (doravante chamado de sistema) desenvolvido em linguagem PHP com o objetivo de produzir relatórios a partir do *thesaurus* empregado. Também apresenta outras rotinas desenvolvidas em linguagem Java para subsidiar os desenvolvimentos da tese, outros *softwares* e bancos de dados utilizados neste trabalho.

O capítulo seis se dedica à apresentação e discussão das configurações e demais detalhes de implementação das duas metodologias de clusterização aplicadas a um conjunto selecionado de traços de personalidade.

O capítulo sete se dedica à apresentação e discussão dos resultados encontrados com as duas metodologias; assim como uma comparação destes dois resultados.

O capítulo oito tece as conclusões acerca dos desenvolvimentos efetuados nesta tese e sugere recomendações para estudos futuros.

2 MÉTODOS QUANTATIVOS NA DETERMINAÇÃO DE CLUSTERS

Neste capítulo se fará uma cobertura geral de diversas famílias de métodos de clusterização disponíveis e conhecidos, tendo como objetivo passar uma visão geral dos métodos existentes. A seguir o mesmo enfocará algoritmos de clusterização passíveis de serem utilizados em problemas de clusterização em grafos, para finalmente focar os dois métodos utilizados neste trabalho redes neurais de Kohonen e algoritmo MCL (Markov Clustering).

2.1 O PROBLEMA GERAL DE CLUSTERIZAÇÃO

O problema geral de clusterização pode ser definido como o problema de se agrupar conjuntos de objetos de forma que, em cada grupo (*cluster*) os objetos são mais similares, segundo um determinado critério a ser especificado, entre si do que quando comparados com quaisquer outros objetos organizados em outros grupos (ou conjuntos).

Pode-se dizer que a análise de *clusters* não é em si mesmo um algoritmo, mas implica em um problema a ser resolvido com informações e critérios adicionais para que estes agrupamentos (*clusters*) de objetos sejam determinados.

Com vista a se determinar estes conjuntos, pode-se utilizar de critérios diferentes entre si segundo suas naturezas:

- Critérios arbitrários: que se valem de entendimentos individualizados, informações e características específicas do problema de clusterização a ser realizado ou mesmo de informações derivadas de áreas específicas do conhecimento (pode-se citar, como exemplo deste último caso, o estudo de ALLPORT e ODBERT (1936), que organizou termos descritivos da personalidade humana em quatro grandes categorias – *clusters* – segundo experiência e informações derivadas da Psicologia).

- Critérios baseados em algoritmos quantitativos: estes critérios baseiam-se na execução de algoritmos que visem a determinar os agrupamentos dos dados iniciais em subconjuntos do mesmo (*clusters*), segundo um conjunto de quesitos pré-especificados. Assim é a lógica inerente a cada um desses algoritmos, conjuntamente com a definição de valores para os parâmetros desses algoritmos que irão determinar estes agrupamentos e, conseqüentemente, delinear as fronteiras de cada *cluster*.

- Critério misto: reunindo execução de algoritmos refinando-se os resultados de um determinado algoritmo de clusterização (e a definição de *clusters*, inicialmente por estes

fornecidas), com informações e conhecimentos adicionais inerentes ao problema de clusterização ou de determinadas áreas do conhecimento.

Esta sessão irá tratar da apresentação de algoritmos quantitativos que podem ser empregados nesta tarefa. Os diversos algoritmos existentes realizam as tarefas de clusterização de dados segundo diferentes noções do que constitui um *cluster* isto é, qual o critério utilizado para a caracterização destes conjuntos e nos métodos utilizados para encontrar estes agrupamentos uma vez definidos estes critérios.

Necessário se faz ressaltar aqui que existem, na literatura, problemas semelhantes e correlacionados aos problemas de clusterização. Um exemplo destes tipos de problemas corresponde aos problemas de classificação. Nesta classe de problemas se deseja organizar os dados em categorias pré-definidas. Os dados classificados em cada uma dessas categorias compartilham uma característica observável em comum.

Exemplos de desenvolvimentos envolvendo tais problemas podem ser encontrados em COLAZO (2009) e PASSOS (2010). O primeiro autor utilizou *Support Vector Machines* (SVM) na implementação de um classificador do risco de morte para pacientes internados com Síndrome Coronariana Aguda. O segundo autor abordou o mesmo problema utilizando Redes Neurais Probabilísticas (PNN).

As noções mais frequentes utilizadas para a definição de um *cluster* incluem:

- grupos de objetos com pequenas distâncias entre si,
- áreas densas do espaço de dados e
- intervalos de distribuições estatísticas particulares.
- regiões de grande concentração de ligações entre vértices de um grafo.

Com base em algumas destas noções diversos algoritmos de clusterização foram desenvolvidos, cada um deles utilizando-se de critérios específicos para definição de um cluster. Estes critérios, por sua vez, baseiam-se nas noções acima mencionadas. Desta forma verifica-se facilmente que, associado a um problema de clusterização tem-se um problema de otimização multi-objetivo.

Mais ainda, o algoritmo de clusterização apropriado, a escolha adequada de sua configuração e a determinação dos seus parâmetros, como por exemplo, a especificação do tipo de função distância ser utilizada, o número esperado de *clusters* etc., depende das características dos dados e dos possíveis usos dos resultados da operação do agrupamento. Deste modo, a análises de *clusters* não deve ser encarada como uma tarefa imediata e

automática (mesmo quando exercida por meio da execução de algoritmos), mas sim como um processo de descobrimento de conhecimento.

Neste sentido, o usuário de algoritmos de clusterização frequentemente se depara com necessidades específicas de pré-processamento de dados e realização de diversas tentativas de clusterização com diferentes configurações de parâmetros até que os resultados das clusterizações alcancem as propriedades desejadas para os *clusters*, na aplicação em pauta.

Alguns dos primeiros trabalhos em análise de *clusters* surgiram na antropologia por DRIVER e KROEBER (1932). Posteriormente trabalhos centrados nestes temas foram aparecendo em outras áreas do conhecimento. Os primeiros trabalhos de clusterização na Psicologia foram apresentados por ZUBIN (1936), por TRYON (1970) e posteriormente CATELL (1943). Este último utilizou técnicas de clusterização de dados para a classificação de traços de personalidade.

De acordo com CASTRO et al (2002) para que um dado apanhado de elementos configure um *cluster* os objetos nele organizados devem possuir, segundo algum critério, maior similaridade entre si do que entre objetos organizados entre diferentes *clusters*; de modo que esta propriedade tem que ser satisfeita em qualquer definição de *cluster*.

Neste sentido, verifica-se que a noção de *cluster* isto é, o conjunto de propriedades necessárias para caracterizar tal arranjo, quando definido por meio de um algoritmo quantitativo, varia de algoritmo para algoritmo; contudo, o resultado destes diferentes algoritmos tem que respeitar a propriedade destacada acima. Com base nestas diferentes propriedades os algoritmos de clusterização podem ser classificados.

Como ressaltado anteriormente, o algoritmo de clusterização mais apropriado para um determinado problema frequentemente precisa ser escolhido experimentalmente, a menos que exista uma razão matemática para se escolher um determinado algoritmo conjuntamente com suas premissas sobre as propriedades que caracterizam um *cluster*.

Em particular, um algoritmo que oferece uma excelente solução de clusterização para um determinado problema pode nem sequer ser utilizável em outros problemas de clusterização. Como nota CASTRO, 2002, algoritmos *K-means* não são capazes de encontrar *clusters* não convexos.

Alguns tipos/famílias de algoritmos de clusterização são apresentados a seguir:

2.1.1 Clusterização hierárquica

Os métodos de clusterização hierárquica são famílias de métodos de clusterização que buscam construir/identificar hierarquias de *clusters*. Estas famílias de métodos se baseiam

em duas estratégias para se obter tais *clusters* hierarquizados; a estratégia aglomerativa e a estratégia divisiva.

Na estratégia aglomerativa, se faz uma abordagem *bottom up*, onde cada observação começa alocada em um *cluster* independente e os pares de *clusters* vão sendo fundidos conforme se move para cima na hierarquia de observações, utilizando-se das ligações existentes entre estas observações.

Na estratégia divisiva, se faz uma abordagem *top down* de clusterização. Nesta estratégia, as observações são inicialmente alocadas em um único *cluster* e vão sendo posteriormente realocadas em *clusters* diferentes (ou seja, vai se fazendo uma divisão do *cluster* inicial em vários *clusters*), baseadas em uma determinada noção de distância.

Nessas famílias de algoritmos, de modo a se decidir que *clusters* devem ser combinados, na estratégia aglomerativa, ou separados na estratégia divisiva é necessária uma medida de dissimilaridade entre conjuntos de observações (que podem ser a distância euclidiana, a distância euclidiana ao quadrado, a distância Manhattan (para mais informações sobre esta medida de distância, vide TAXICAB METRIC, 2012), a distância Mahalanobis (apresentada em MAHALANOBIS METRIC, 2012) dentre outras.

Isso é possível por meio de uma medida de distância e por um critério de ligação, que especifica a dissimilaridade de conjuntos como função de distâncias entre pares de observações entre conjuntos.

Alguns desses algoritmos de clusterização aglomerativa são: SLINK (SINGLE-LINK EXAMPLE, 2012), CLINK (vide SINGLE LINK, COMPLETE-LINK & AVERAGE-LINK CLUSTERING, 2012), UPGMA (*Unweighted Pair Method with Arithmetic Mean*, cuja apresentação geral está disponível em UPGMA, 2012), UPGMC etc. (UPGMA, 2012).

De uma forma geral, os algoritmos aglomerativos tem complexidade $O(n^3)$ o que os tornam inadequados para grandes quantidades de dados e os algoritmos divisivos possuem complexidade $O(2^n)$ sendo ainda menos apropriados para grandes quantidades de dados.

As hierarquias entre *clusters* podem ser organizadas e visualizadas em dendogramas; que são os instrumentos mais utilizados para estes propósitos.

A seguir se apresenta diferentes medidas de distância comumente utilizadas por estes algoritmos.

2.1.1.1 Medidas de distância utilizadas por diferentes algoritmos de clusterização

Algoritmos de clusterização que utilizam cálculos de distâncias entre elementos como etapa para a definição de *clusters* podem se valer de diferentes medidas de distância. Dentre as principais medidas destacam-se:

- Distância Euclidiana: representada pela norma do vetor diferença entre dois vetores de dados, a e b: $\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2}$

- Distância Euclidiana ao Quadrado: representada pelo quadrado da distância euclidiana entre dois vetores de dados, a e b: $\|a - b\|_2^2 = \sum_i (a_i - b_i)^2$

- Distância Manhattan: representada pelo somatório dos módulos das diferenças das coordenadas dos vetores a e b: $\|a - b\|_1 = \sum_i |a_i - b_i|$

- Distância Similaridade Cosseno: dada por: $\frac{a \cdot b}{\|a\| \|b\|}$

2.1.1.2 Critérios de ligação comumente adotados por algoritmos de clusterização hierárquica

Com relação aos critérios de ligação, os mais comumente utilizados em análise de *clusters* são:

- Ligação Completa ou Máxima (*Complete Linkage*): a ligação completa, ou máxima entre dois *clusters*, X e Y, é dada por: $\max_{x \in X, y \in Y} d(x, y)$. Esta ligação é utilizada em clusterização hierárquica aglomerativa, onde a distância entre dois *clusters* é dada pela distância entre o par de elementos (cada um pertencendo a um *cluster* diferente) mais distantes entre si.

- Ligação Mínima (ou *Single Linkage*): é a distância entre dois clusters X e Y, definida por: $\min_{x \in X, y \in Y} d(x, y)$. Com esta ligação, a distância entre dois *clusters* é dada pela distância entre o par de elementos, cada um pertencendo a um *cluster* diferente mais próximo entre si.

- Ligação Média (*Average Linkage*): define a distância entre dois clusters X e Y por: $\frac{1}{|X||Y|} \sum_{x \in X} \sum_{y \in Y} d(x, y)$. Com esta ligação, a distância entre dois *clusters* é dada pela distância

entre o par de elementos, cada um pertencendo a um *cluster* diferente mais próximo entre si. Utilizando-se esta ligação a distância entre dois *clusters* é definida como a média das distâncias entre todos os pares de elementos de cada *cluster*. Esse critério de ligação é muito usado em algoritmos de clusterização hierárquica; especialmente no algoritmo hierárquico

bottom up conhecido como UPGMA (*Unweighted Pair Group Method With Arithmetic Mean*).

- Ligação Baseada em Centróide: este critério de ligação entre dois *clusters* X e Y é definido como por: $\|\varphi_X - \varphi_Y\|_2$, onde φ_X e φ_Y são os centróides dos *clusters* X e Y respectivamente, ambos obtidos pela média aritmética de seus respectivos elementos. É utilizado no algoritmo de clusterização aglomerativa conhecido como UPGMC (*Unweighted Pair-Group Method using Centroids*). Neste algoritmo, a distância entre dois *clusters* é computada utilizando-se somente a distância Euclidiana. Isto é, as seguintes condições devem ser respeitadas:

$$1) d(X, Y) = \|\varphi_X - \varphi_Y\|_2$$

e

$$2) \varphi_X = \frac{1}{n_X} \sum_{i=1}^{n_X} X_i$$

- Método de Ward: ao se adotar este método como critério de ligação, deve-se especificar uma função objetivo a ser otimizada durante a escolha de que *clusters* serão fundidos, em um problema de clusterização aglomerativa.

2.1.1.3 Outros algoritmos aglomerativos/divisivos de clusterização

Algoritmos aglomerativos/divisivos de clusterização baseados em classificadores de Bayes e modelos de mistura também tem sido propostos na literatura. Um exemplo de tal algoritmo é o proposto por BAKER e MACCALLUM (1998); que programaram um método de classificação de palavras, a partir de documentos de texto, baseado em um classificador de Bayes onde se supunha que os dados estavam sendo gerados por modelos de misturas de densidades paramétricas.

2.1.2 Clusterização baseada em centróide (algoritmo *K-means*)

Na clusterização baseada em centróides, os *clusters* são identificados com base em elementos centrais, localizados no meio dos dados a serem clusterizados; utilizando estes elementos como referência na determinação dos agrupamentos (por exemplo, utilizando a distância entre elementos centrais nesta determinação).

No algoritmo *K-means* (algoritmo que segue esta filosofia de clusterização) define-se a priori o número de *clusters* (diga-se k) que se deseja definir e encontram-se os k *clusters* e se aloca cada um dos dados ao *cluster* que possui seu centro mais próximo desse dado, de tal forma que se minimiza o quadrado das distâncias nesse processo. Com o número de *clusters*

pré-definido o processo de clusterização por este algoritmo corresponde a um problema de otimização NP-árduo. Este algoritmo repetidamente encontra o centroide de cada conjunto na partição e então reparticiona os dados de acordo com o centroide mais próximo para cada observação, para um conjunto discreto de pontos. Este processo é repetido até a convergência do algoritmo.

O algoritmo *K-means* possui diversas variantes, como o algoritmo de Lloyd, onde o espaço de dados é contínuo, o algoritmo *k-medoids*, onde se restringe o centroide a ser um elemento do conjunto de dados, *K-medians*, o centroide é a mediana do conjunto de dados, versões *fuzzy* do *K-means*, etc.

2.1.3 Clusterização baseada em densidades

Nesta filosofia de clusterização os *clusters* são definidos como áreas de maior densidade do que o restante do conjunto de dados. Dados localizados entre áreas de grandes densidades são considerados ruído ou pontos de fronteiras entre *clusters*.

O algoritmo mais conhecido nesta tradição é o DBSCAN (*Density-Based Algorithm for Discovering Clusters in large Spatial Databases with Noise*) de ESTER, KRIEGEL, SANDER e XU (1996).

Outro algoritmo tradicional nesta tradição de clusterização OPTICS (*Ordering Points to Identify the Clustering Structure*), de ANKERST, BREUNIG, KRIEGEL e SANDER (1999).

2.1.4 Métodos de clusterização de dados baseados em análise fatorial

Dentre os métodos de clusterização baseados em análise fatorial pode-se citar o apresentado por TRYON E BAILEY (1970) em 1939, considerado um dos trabalhos pioneiros em análise de *clusters*. Seguindo esta tradição, CATELL (1943) determinou *clusters* de traços de personalidade. Este último autor utilizou-se de coeficientes de correlação da análise fatorial em respostas de questionários para a determinação de *clusters* de traços.

2.1.5 Outros métodos de clusterização

Outros algoritmos de clusterização também tem se consagrado na literatura a cerca do tema. Dentre eles pode-se citar:

- O algoritmo CLARANS, de RAYMOND e JIAWEI (1994), baseado em busca aleatória e que oferece uma solução de clusterização para dados espaciais.
- Como já apresentado anteriormente, ZUBIN (1938) apresentou um método de clusterização, de indivíduos, quanto a um determinado traço de personalidade pré-

especificado, baseados no grau de concordância de *scores* obtidos em respostas a quesitos em questionários. Este trabalho de análise de *clusters* foi pioneiro como método de clusterização na Psicologia.

2.2 ALGORITMOS DE CLUSTERIZAÇÃO DE GRAFOS

Uma parte importante dos algoritmos de clusterização se atém ao problema de clusterização em grafos. Este problema se resume em se organizar os vértices de um grafo segundo um determinado critério, não se relacionando necessariamente com problema diferente do problema de clusterização de pontos/vetores (*vector clustering*) em um determinado espaço, uma vez que o critério utilizado para se definir clusters em grafos pode ser diferente ao critério de se definir clusters baseados nas distâncias entre vetores.

De forma genérica pode-se definir um grafo da seguinte forma: um grafo $G(V, A)$ é definido pelo par de conjuntos V e A ; onde V é o conjunto não vazio dos vértices do grafo e A é o conjunto de pares ordenados $a = (v, w), v, w \in V$ definindo as arestas (ou arcos) do grafo.

Por exemplo, pode-se definir como critério na definição de tais agrupamentos em grafos a densidade de ligações entre vértices; onde as distâncias entre vértices do grafo não são utilizadas nessas definições (como no algoritmo MCL, desenvolvido por DONGEN, 2000).

A figura1 abaixo apresenta esquemas gráficos genéricos de clusterização de vetores e de clusterização em grafos.

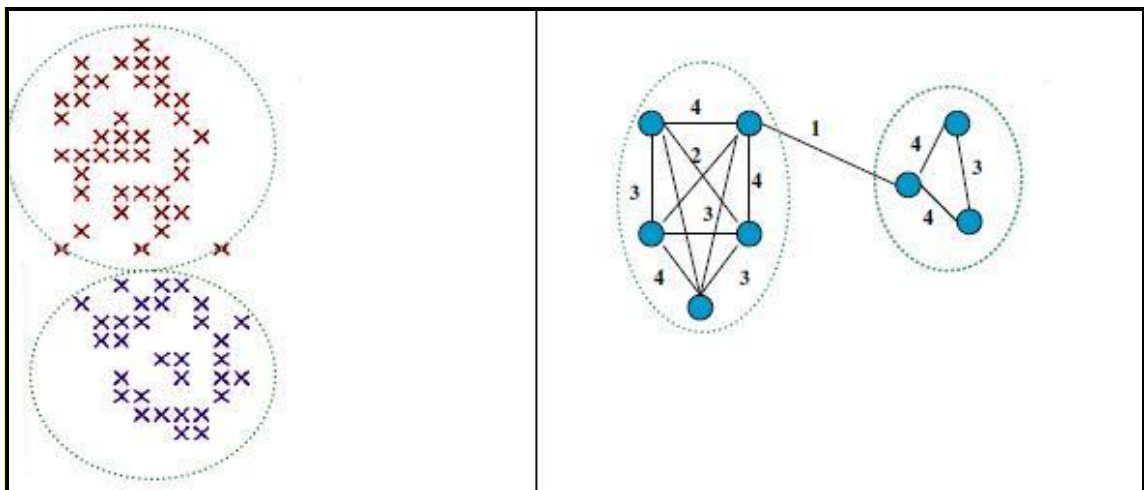


Figura1: Esquemas gráficos de clusterização de vetores e em grafos.

Fonte: MACROPOL, 2009, p. 4.

Na figura 1 acima, os *clusters* de vetores (à esquerda) foram definidos com base nas coordenadas dos pontos e os *clusters* no grafo apresentado à direita se basearam no critério de densidade de ligações entre vértices.

O problema de clusterização em grafos, de acordo com GERASOULIS E YANG (1992) pode ser encarado como um problema de mapeamento (*mapping*) dos vértices de um grafo em *clusters* que podem ser rotulados por um determinado *label*.

Este tipo de problema possui elevada complexidade algorítmica, sendo NP-COMPLETO para um grafo geral e para diversas funções custo.

SARKAR (1989), CHRETIENNE (1989), e PAPADIMITRIOU (1990) demonstraram que, problemas envolvendo aplicações deste tipo na determinação de arquiteturas ótimas de computação paralela, onde a função custo é a minimização do tempo de processamento em paralelo numa arquitetura virtual completamente conectada, sem um número máximo de processadores são NP-ÁRDUO (NP-HARD).

Um dos problemas abordados em clusterização de grafos é o problema da determinação de *clusters* em grafos acíclicos. Este problema também possui uma série de aplicações (como por exemplo, na partição de circuitos, como apresentado por WONG, YOUNG e MAK, 2003).

Deste modo, se vê que o problema da determinação de *clusters* em grafos é um problema muito importante tanto na Ciência da Computação, como na Pesquisa Operacional e em diversas outras áreas do conhecimento; sendo também um campo bastante suscetível a contribuições algorítmicas.

No entanto, apesar de todos os desenvolvimentos feitos nesta área, não se verificou, até o momento, nenhum algoritmo capaz de fornecer uma representação de cada vértice do grafo para que pudessem ser utilizados por determinados algoritmos de clusterização, em particular rede neurais de Kohonen e *K-means*, em problemas onde não se dispusesse de nenhuma informação adicional a cerca do grafo que não as adjacências entre estes mesmos vértices.

Para que tal possibilidade se concretizasse, se fazia necessário, ser possível se calcular diretamente a distância euclidiana entre cada par de vértices, por meio dessa codificação e que esta codificação transformasse cada vértice em uma estrutura tal que pudesse ser utilizada por estes algoritmos.

A representação dos vértices de um grafo como um vetor possibilitaria a sua imediata utilização em diversas máquinas de aprendizado e algoritmos que utilizem como *input* esse tipo de estrutura, como por exemplo, redes SOM e algoritmos *K-means*.

Este capítulo tratará da metodologia de codificação proposta para tais finalidades, em grafos acíclicos assim como de outros métodos que podem ser empregados para tais propósitos.

2.2.1 Aprendizado não supervisionado em problemas de clusterização

Os problemas de clusterização podem empregar máquinas de aprendizado nas suas soluções, tais como redes neurais.

Formalmente, dividem-se os treinamentos dessas redes em treinamentos supervisionados e não supervisionados.

Nos treinamentos supervisionados, tanto os dados de entrada como as saídas desejadas são apresentadas à rede, de forma que uma medida de erro possa ser computada ao longo desse treinamento. Com base nessa medida de erro os algoritmos de treinamento da rede vão ajustando seus pesos sinápticos de forma a melhorar seu desempenho.

Esta tese irá utilizar uma rede neural com treinamento não supervisionado para as tarefas de clusterização de traços de personalidade a partir da estrutura de relacionamentos entre esses traços em um *thesaurus*.

Mais particularmente se irá utilizar uma rede SOM (*Self Organizing Map*).

2.2.1.1 Redes neurais artificiais de Kohonen (ou SOM)

A rede neural de Kohonen é um tipo de rede neural que realiza treinamento não supervisionado e pode de uma forma geral resolver problemas de clusterização e de reconhecimento de padrões.

Essas redes neurais implementam um mapeamento de um espaço de dados de entrada (de alta dimensão) em uma grade (*grid*) regular de baixa dimensão enquanto preserva as relações topológicas entre os dados. Esta grade (ou mapa) consiste de um conjunto de nós (ou vértices), com um vetor de pesos $m_i \in \mathbb{R}^n$ associado a cada nó da grade.

Esta grade também pode ter diferentes formatos. Os tipos mais comuns são: retangular, hexagonal e irregular, neste último caso, os espaçamentos e organização dos nós na grade são irregulares; o que implica também que as ligações entre os nós da grade variem de forma irregular. De acordo com KOHONEN (2001) a arquitetura hexagonal é mais efetiva para visualização dos dados (e dos *clusters*). As figuras 2 e 3 apresentam alguns exemplos de arquiteturas de mapas auto-organizáveis.

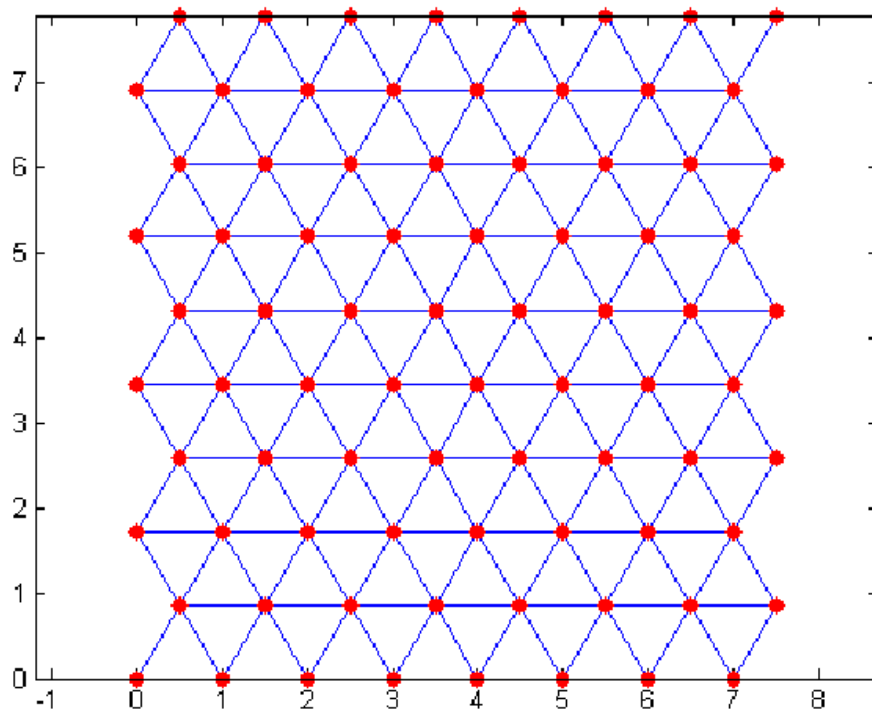


Figura 2: Exemplo de topologia hexagonal de um mapa auto-organizável

Fonte: BEALE, HAGAN e DEMUTH (2014)

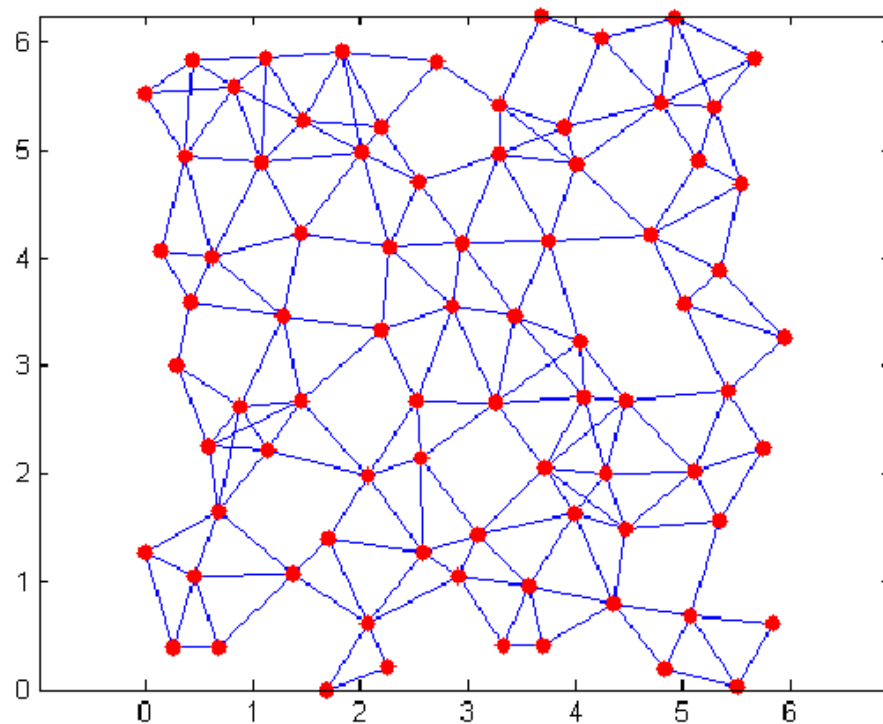


Figura3: Exemplo de topologia aleatória de um mapa auto-organizável

Fonte: BEALE, HAGAN e DEMUTH (2014).

As redes SOM proveem um mapeamento de dados de alta dimensão em espaços de baixa dimensão (frequentemente de duas dimensões). Esse mapeamento preserva a topologia dos vetores.

Esse mapa é chamado de mapa topológico e preserva os relacionamentos entre estes vetores do espaço de entrada (*input*) da rede. Como se verá, ele é responsável pela clusterização de dados de entrada na rede neural, agrupando dados similares em nós em uma mesma vizinhança.

Para tanto, cada dado de entrada x_i é comparado com cada elemento-modelo que a rede neural inicialmente se utiliza como candidato à representação deste dado, a cada modelo m_i corresponde um nó da grade de baixa dimensão. Cada um desses nós é chamado na literatura de *codebook vector*. A figura abaixo apresenta um esquema dessa classificação inicial dos dados de entrada por meio das distâncias com cada elemento-modelo.

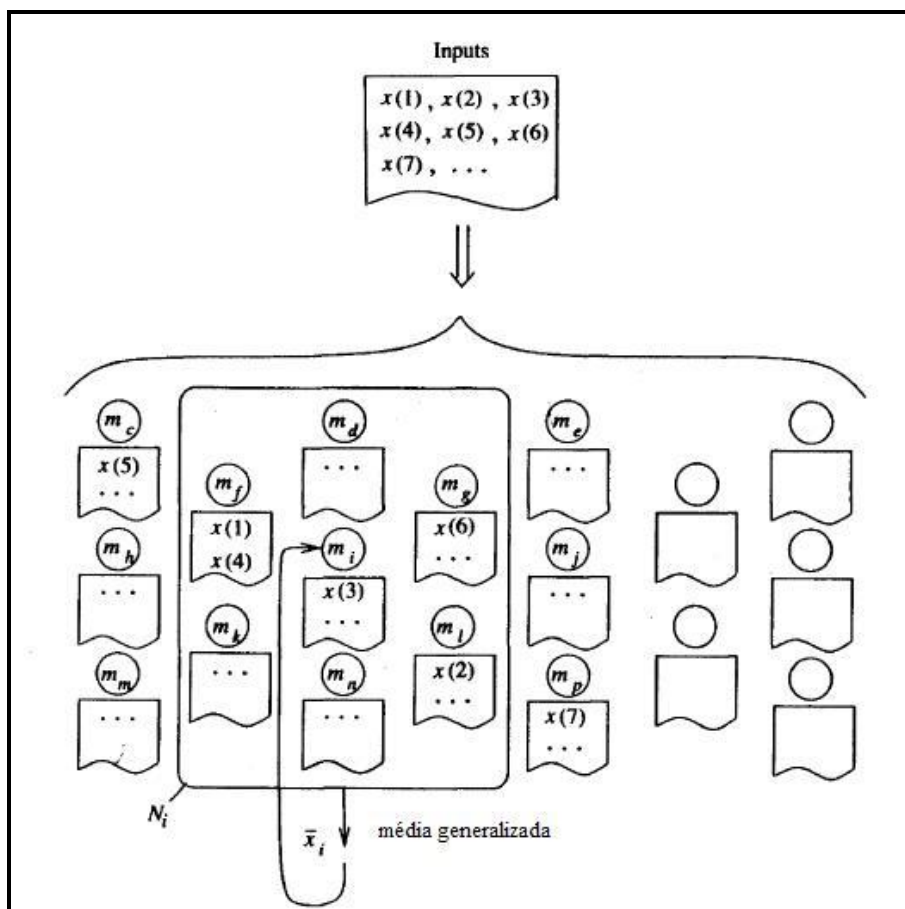


Figura4: Exemplo de identificação inicial de cada dado de entrada com os modelos que melhor os representam.

Fonte: KOHONEN (2001).

O modelo que melhor representa x_i é escolhido pelo critério de *Best Matching Node* (BMN) a partir da utilização de determinada medida de distância.

Assim, considerando-se a distância Euclidiana, o modelo m_c que representará x_t será, considerando-se nesta seção, daqui para frente para fins expositivos, que cada dado de entrada x_t e cada modelo m_i são vetores reais; isto é, que $x_t \in \mathfrak{R}^n$ e $m_i \in \mathfrak{R}^n$ dado por:

$$c = \underset{i}{\operatorname{argmin}} \left\| \frac{1}{\sqrt{N}} x_t - m_i \right\|$$

Para cada dado de entrada apresentado à rede, se irá identificar o modelo $m_i \in \mathfrak{R}^n$ que melhor se aproxima deste dado, segundo uma determinada medida de distância.

Durante o processo de aprendizagem, os nós na grade – representando os modelos disponíveis, que são topograficamente próximos entre si irão ativar uns aos outros para aprender algo sobre o mesmo dado de entrada x_t . Para tanto, no entorno de cada nó na grade, representando um modelo m_i , se definirá uma vizinhança N_i .

Assim, a rede neural alterna uma etapa de comparação das entradas x_t com cada modelo disponível segundo o critério *Best Matching Node* (BMN), com uma etapa de atualização de cada modelo m_i . Este processo leva a um ordenamento dos vetores de entrada no mapa segundo a similaridade entre eles e os nós da grade representando os modelos, que são atualizados para levarem em consideração informações dos dados de entrada a cada iteração.

Durante a fase de treinamento os dados de entrada são ordenados na grade de tal forma que dados similares entre si vão ficando próximos uns dos outros e dados menos similares vão sendo localizados em regiões mais afastadas da grade.

Como já ressaltado, o grau de similaridade entre os dados de entrada da rede é medido indiretamente, a partir das distâncias de cada um com cada modelo e suas posteriores atualizações que a rede utiliza.

No mapa topológico resultante os relacionamentos (similaridades) entre os dados de entrada - e entre estes e os nós da grade - são preservados e apresentados em dimensão baixa para imediata visualização. Estes dados só assim clusterizados em função de suas similaridades, segundo suas distâncias aos modelos obtidos ao final de todas as iterações.

KOHONEN (2001, p. 111) sugere que esta atualização dos modelos seja genericamente dada por:

$$m_i^{(t+1)} = m_i + h_{ci} \frac{1}{N} x_t - m_i$$

Onde:

t : é um índice inteiro, representando momentos discretos do tempo,

$m_i(t)$: é o modelo m_i no instante (ou iteração) t ;

$m_i(t+1)$: é o modelo m_i no instante (ou iteração) $t+1$;

$h_{ci}(t; m_i(t))$: é uma função *kernel* avaliada em $x(t; m_i(t))$. Essa função é monotonamente decrescente, o que implica que quanto maior a distância de $x(t)$ a $m_i(t)$, menor será seu valor e conseqüentemente, menos informação de $x(t)$ será levada em consideração na atualização de m_i . No processo de relaxação acima citado, essa função age como uma função vizinhança, sobre os pontos do mapa.

Para a convergência do processo, é necessário que $h_{ci} \rightarrow 0$ quando $t \rightarrow \infty$.

Frequentemente se adota como *kernel* h_{ci} a seguinte função:

$$h_{ci}(t) = h(\|r_c - r_i\|, t) = \alpha(t) \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right)$$

Onde:

$r_c, r_i \in \mathbb{R}^2$ são os vetores de localização dos nós c e i , na grade, respectivamente.

A função $h_{ci}(t)$ define uma vizinhança de pontos na grade, em torno do nó c . Assim, $h_{ci}(t) = 0$ se o nó $i \notin N_c$.

$\alpha(t)$ é a taxa de aprendizagem. Tanto $\alpha(t)$ como $N_c(t)$ são monotonicamente decrescentes no tempo. Finalmente, $\sigma(t)$ (também monotonicamente decrescente) define a amplitude de valores do *kernel*.

Demais características da rede neural de Kohonen:

- nenhum conhecimento ou hipótese sobre a distribuição dos dados para o caso de dados aleatórios de entrada é necessário para se utilizar a rede SOM.
- o algoritmo de aprendizagem é simples de ser compreendido e utilizado,
- não há necessidade de nenhum sinal (ou série de dados) para guiar o processo de treinamento,
- a similaridade entre dados pode ser facilmente encontrados em dados de alta dimensão e dados similares são mapeados em regiões de vizinhança de baixa dimensão, facilitando a visualização dos *clusters* e,
- a busca pelo melhor nó representando cada dado (BMN) de entrada é exaustiva e computacionalmente cara para dados de alta dimensionalidade ou mapas grandes.

Em particular, KOHONEN (2001) sugere que a função kernel a ser utilizada seja dada por:

Em essência, essas redes neurais funcionam como um método de redução de dimensionalidade dos dados, e de agrupamento dos mesmos em categorias (ou classes) por similaridade.

Apresentações que cobrem com profundidade as redes SOM podem ser encontradas em WASSERMAN (1993), PATTERSON (1996), KOHONEN (2001), RIPLEY (2004), CALÔBA (2009) e PEREIRA E RAO (2009).

2.2.1.2 Redes SOM em clusterização de árvores e florestas

No esforço de se aplicar uma rede neural do tipo SOM no problema de clusterização de traços de personalidade tratado nesta tese, a partir do TTT adquirido, verificou-se inicialmente que nenhum dos métodos de codificação das palavras presentes no *thesaurus*, propostos na literatura para a criação de dados de *input* para esta rede neural, se aplicava ao problema em questão.

Neste problema, praticamente toda a informação que se dispunha a cerca dos grafos (representando as relações entre palavras selecionadas no *thesaurus*) onde se pretende encontrar *clusters*, se referia à adjacência entre os vértices (representando as palavras neste repositório) dos mesmos.

Com basicamente somente estas informações a cerca destes grafos não se podia utilizar as redes neurais SOM; sendo necessária como etapa anterior, desenvolver uma forma de codificação (ou de representação) dos vértices desses grafos para que pudessem ser utilizados neste tipo de rede neural.

Tuevo Kohonen (2001) chega a propor também uma medida de distância entre palavras diferentes baseada nas diferenças das sequências de caracteres que formam as palavras. Contudo, esta medida de distância não se aplica ao problema aqui tratado, já que palavras diretamente relacionadas no TTT podem ser formadas por sequências de caracteres totalmente diferentes, uma vez que o que as une não é a similaridade entre elas nas sequências desses caracteres, mas sim nos sentidos de associação entre as mesmas.

Assim, estas noções de distância não têm utilidade para o problema de clusterização em pauta, uma vez que palavras formadas por sequências similares de letras podiam não estar associadas umas às outras em nenhum sentido de associação identificado no *thesaurus* utilizado.

De forma análoga, verificam-se quase sempre na própria estrutura de relações entre estas palavras no TTT que, palavras muito diferentes, em termos das sequências de letras que as formam frequentemente estavam associadas diretamente umas com as outras, segundo um determinado sentido de associação.

Assim, os graus de similaridade na sequência de caracteres que formam as palavras pouco ou nada auxiliam na identificação do grau de proximidade entre palavras distintas, no problema em pauta. Necessitava-se, deste modo, de uma metodologia que pudesse identificar o grau de proximidade destas palavras, no *thesaurus* utilizado, onde nada mais se dispunha além das relações de adjacência entre elas.

Deve-se destacar aqui que embora a literatura sobre redes SOM apresente versões modificadas deste tipo de rede para dar conta de particularidades de dados a serem clusterizados (HAGENBUCHNER ET al, 2003 e KOFI et al, 2009) ou para melhorar a desempenho computacional no treinamento da mesma (como em KUN-SEOK ET al, 2001), não se encontrou na literatura nenhuma forma de se adaptar a rede de Kohonen às particularidades do problema aqui abordado.

Fazia-se necessário, portanto, uma metodologia que levasse em consideração as relações de adjacência entre estas palavras e que as traduzisse em uma medida numérica de distância entre elas.

Tendo em vista estas propriedades desejáveis, pensou-se em representar estas palavras, (ou vértices do grafo representando as associações entre elas), como vetores, segundo o algoritmo a seguir apresentado.

Esta metodologia proposta de codificação dos vértices de um grafo como vetores, se adéqua bem a problemas onde os vértices não possuam coordenadas explícitas – ou conhecidas a priori - e onde tudo o que se disponha sejam as relações de adjacência entre eles.

Tais considerações forneceram motivação para o desenvolvimento do algoritmo de codificação de vértices de um grafo, exposto a seguir.

2.2.1.3 Metodologia proposta para codificação dos vértices de um grafo visando-se aplicações de clusterização

No esforço de se tentar resolver este problema, para que se pudessem utilizar redes SOM na determinação de *clusters* nos grafos considerados, conseguiu-se desenvolver, de forma inédita, um algoritmo que codifica os vértices de um grafo como vetores para que possam servir de dados de entrada para estas redes. Não obstante o algoritmo ainda não ter

sido testado desta forma, o algoritmo proposto pode ser utilizado também para produção de *inputs* a algoritmos do tipo *K-means* (não utilizados neste trabalho) e de outros algoritmos que promovam clusterização de dados utilizando medidas de distância.

Como será apresentado, embora este método de codificação tenha sido concebido para aplicações em grafos acíclicos, a experiência obtida até o momento da finalização desta tese, com grafos cíclicos foi muito positiva no que se refere aos resultados obtidos em aplicações de clusterização.

Quando a metodologia é aplicada a grafos acíclicos, o quadrado das distâncias euclidianas entre vetores (representando os vértices do grafo sendo clusterizado) de uma mesma componente conexa representa o número mínimo de passos necessários para se chegar a um vértice do grafo partindo-se de outro.

Na metodologia desenvolvida, cada vértice do grafo vai entrar em uma lista de vértices em processo de codificação (LVPC). As palavras entram nesta lista, uma por vez.

Cada vértice que entra na lista de codificação será representado por um vetor composto em sua quase totalidade por valores nulos ou unitários. A dimensão dos vetores que representam os vértices já incluídos na LVPC vai aumentando na medida em que novos vértices entram nesta lista. Assim, a dimensão desses vetores fica em aberto (e vai aumentando gradualmente) até que o último vértice do grafo seja incluído na lista.

Cada vértice que entrar nesta lista será representado por uma sequência de números separados (no exemplo) por vírgulas, correspondentes às coordenadas do vetor que representará de forma única este vértice.

A dimensão desses vetores não ficará totalmente determinada até que o algoritmo chegue ao fim, o que se dará quando todos os vértices tiverem entrado nesta lista.

Como se verá, cada vetor representando um determinado vértice que acabou de entrar na LVPC imediatamente repete a estrutura de outro vetor que está lhe servindo de referência na sua codificação (que por sua vez representa o vértice de referência nesta etapa do processo de codificação, como será visto mais adiante), repetindo todas as suas coordenadas até o momento da sua entrada na LVPC. Chamaremos este último vetor de vetor de referência dessa etapa do processo de codificação (VREPC).

Após esta repetição, o primeiro vetor representando o nó que acabou de entrar na lista, terá sua dimensão aumentada em uma unidade. Isto é feito através da adição de mais uma coordenada neste vetor (também de valor unitário).

O aumento dimensional efetuado neste vetor é seguido por aumentos unitários de dimensão nos demais vetores da LVPC; o que por sua vez é efetuado por meio da adição de coordenadas nulas nesses vetores da lista.

Deste modo, cada vetor representando um vértice que acabou de ser incluído na LVPC repete a estrutura do seu vetor de referência neste momento do processo de codificação e tem sua dimensão aumentada (sempre através da inclusão de uma coordenada unitária). Toda vez que isto ocorre, esta ação é sempre seguida por um procedimento similar de aumento de dimensão nos demais vetores da lista (só que nestes casos, através da inclusão de coordenadas nulas nos mesmos).

Após a inicialização do algoritmo, o vértice inicialmente escolhido para ser codificado entrará na LVPC e servirá como referência para a codificação dos vértices do grafo vizinhos a ele (isto é, dos vértices que estão diretamente ligados a ele).

Após a inclusão na LVPC do vértice escolhido (que se chamará aqui, para fins didáticos, de vértice A) o algoritmo procede na identificação e inclusão (uma por vez) de todos os seus vértices vizinhos (que se chamará de A-vizinhos) na lista.

Uma vez que todos os A-vizinhos tenham entrado na lista o algoritmo procede à codificação de todos os vértices vizinhos dos A-vizinhos (já na LVPC), procedendo, para tanto, à identificação e colocação na LVPC dos vértices adjacentes aos vértices A-vizinhos. Isto é feito da seguinte forma:

- 1) Seleciona-se um vértice A-vizinho já incluso na LVPC. Este vértice servirá de vértice de referência na codificação de todos os vértices a ele adjacentes (e não ainda incluídos na lista). Identificam-se todos estes vértices.
- 2) Seleciona-se e inclui-se (um por vez) na LVPC cada um desses vértices adjacentes a esse VREPC (que na etapa atual do algoritmo é o A-vizinho selecionado em 1), até que todos eles tenham entrado na lista.
- 3) Uma vez que todos os vértices adjacentes ao A-vizinho selecionado tenham entrado na lista, outro A-vizinho é selecionado na LVPC (onde essas escolhas dos A-vizinhos devem respeitar a ordem nas quais estes entraram na LVPC).

As etapas 1, 2 e 3 são repetidas até que todos os vértices adjacentes a todos os A-vizinhos tenham entrado na LVPC (e, portanto, tenham começado a ser codificados como vetores). Pode-se constatar que a execução dessas três etapas envolve recursividade.

A figura 5 abaixo provê um esquema ilustrativo das etapas 1), 2) e 3).

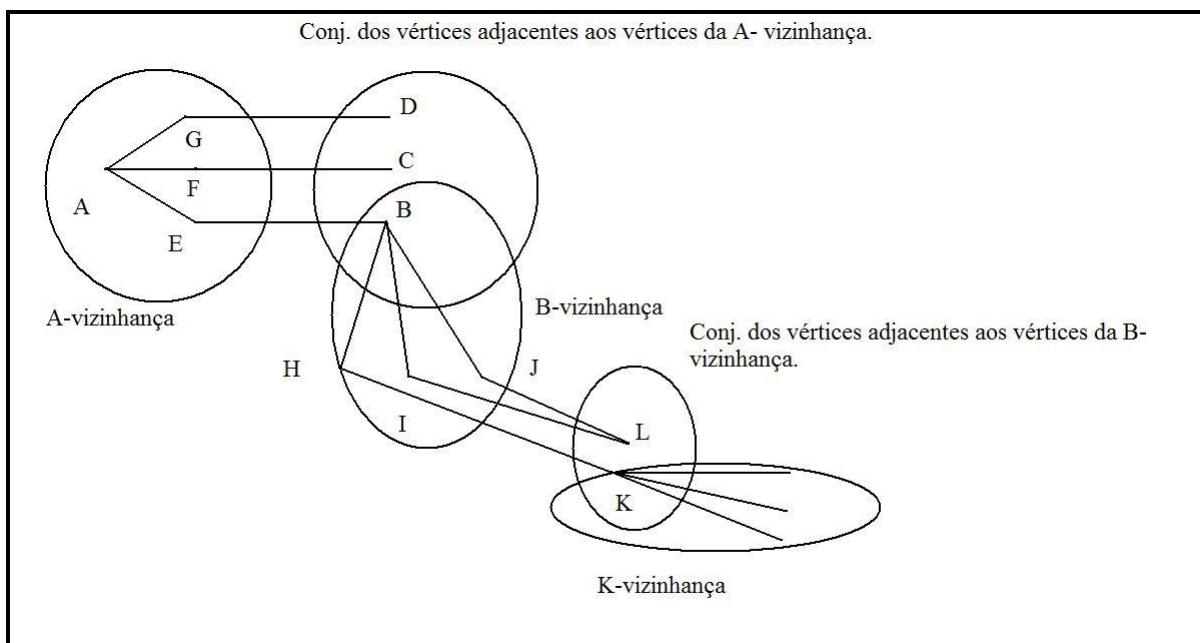


Figura 5: Esquema ilustrativo das etapas 1), 2) e 3) do algoritmo de codificação

A figura acima ilustra as etapas 1, 2 e 3 do algoritmo de codificação. Inicialmente escolhe-se um vértice inicial para inicialização do algoritmo. Escolhido o vértice A para ir para LVPC, essa escolha define automaticamente a A-vizinhança, definida pelos vértices diretamente ligados ao vértice A. Depois que todos os vértices da A-vizinhança vão para a LVPC o algoritmo prossegue identificando todos os vértices adjacentes aos vértices da A-vizinhança, os enviando para a LVPC.

Estas etapas são repetidas até que todos os vértices do grafo sejam enviados para a lista de codificação. Neste sentido, identifica-se o vértice B e em torno dele define-se a B-vizinhança. Após todos os vértices da B-vizinhança ter sido enviados a LVPC os vértices adjacentes a estes últimos são identificados e enviados à LVPC. Finalmente, neste último conjunto identificam-se os vértices vizinhos de K, e assim definindo-se a K vizinhança.

Após a codificação de todos os vértices da K-vizinhança o algoritmo se encerra.

Uma vez incluídos todos estes vértices, o algoritmo então prossegue, da mesma forma, identificando (e adicionando à lista) os vizinhos dos últimos vértices a entrarem na LVPC e os vértices adjacentes a estes últimos, até que todos os vértices do grafo (ou do componente conexo do mesmo onde a codificação estava se dando) tenham entrado na lista.

No caso de um grafo conexo, após o último vértice ter entrado na lista, a codificação acaba. No caso de um grafo desconexo, após a entrada do último vértice a ele pertencente, na LVPC, o algoritmo elege outro componente conexo (como é o caso do exemplo a ser mostrado mais adiante) para a continuidade do processo de codificação.

Neste ultimo caso (grafo desconexo), uma vez que todos os vértices de uma dada componente sejam incluídos na lista, deve-se guardar a informação de que aqueles nós pertencem àquela componente específica e que os próximos vértices a entrarem na lista

pertencerão à outra componente. Uma vez que esses novos vértices comecem a entrar na lista, a lógica do algoritmo é repetida.

Entretanto, ainda nos casos de grafos desconexos, após a inclusão de todos os vértices do grafo na lista o algoritmo incluirá, em cada um dos vetores, coordenadas adicionais, visando à atribuição de distâncias explosivas (muito grandes) entre vetores pertencentes a diferentes componentes conexas.

Esse último procedimento visa garantir que as distâncias euclidianas computadas entre vetores representando vértices pertencentes a diferentes componentes do grafo sejam desproporcionalmente maiores do que as distâncias atribuídas entre vetores em uma mesma componente (isto será importante na utilização de vetores assim construídos em problemas de clusterização de grafos que usem tais estruturas).

O algoritmo a ser detalhado por meio de um exemplo tem o mérito de proporcionar uma forma simples e sistemática de representar cada vértice do grafo por meio de uma estrutura que possa servir diretamente de *input* a redes neurais do tipo SOM (embora, essas estruturas podem ser utilizadas como *input* para outros algoritmos de clusterização).

O algoritmo foi concebido para ser utilizado em grafos não direcionados com arestas não valoradas.

No caso de grafos cíclicos, o algoritmo de codificação fornece aproximações das distâncias mínimas entre os vértices do grafo que participam de ciclos. Os experimentos efetuados com esta metodologia se centraram em torno da distância euclidiana e do quadrado dessa distância; a partir das quais a metodologia foi analisada.

Como já mencionado, e como será apresentado no capítulo 7, obteve-se excelentes resultados com este algoritmo em problemas de clusterização de palavras provindas do TTT, no sentido de que a alocação fornecida pela rede neural SOM, utilizando os vetores construídos pelo algoritmo propiciaram alocações consistentes com as estruturas de ligações presentes entre os vértices dos grafos utilizados nesses experimentos.

Este algoritmo será apresentado por meio de um exemplo:

A figura 6 abaixo apresenta um grafo (não conexo), composto por duas componentes conexas.

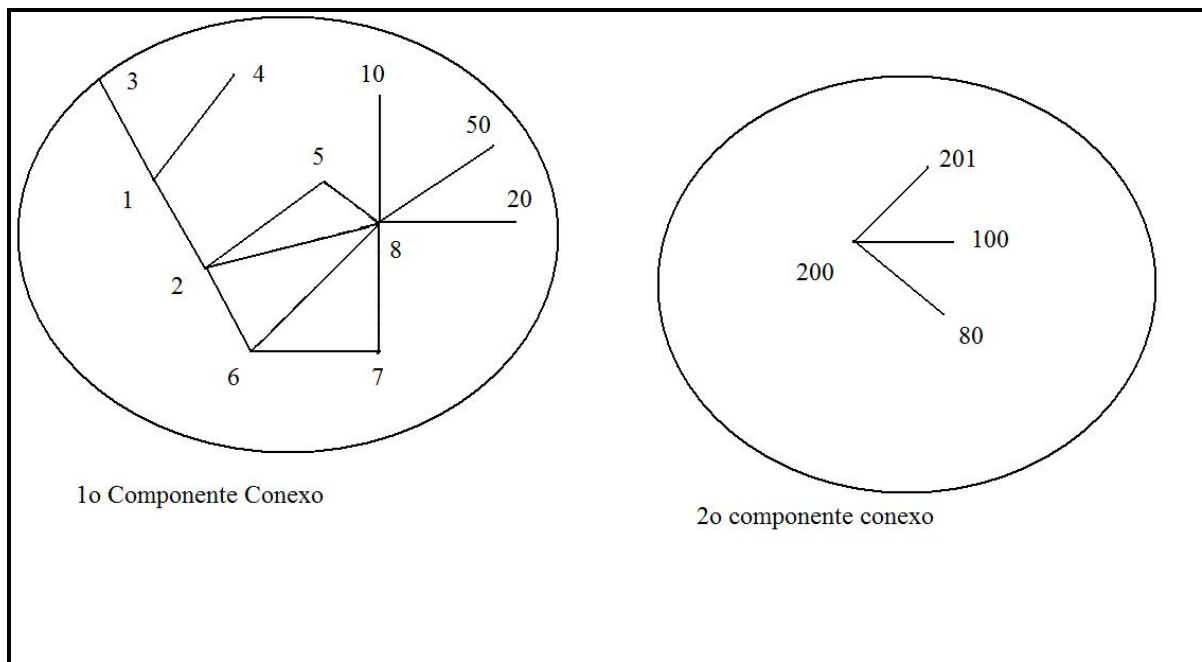


Figura 6: Grafo com duas componentes conexas, utilizado de exemplo do algoritmo de codificação dos vértices de um grafo.

Em Teoria de Grafos, uma componente conexa de um grafo não direcionado (como o do exemplo) é um subgrafo no qual quaisquer dois vértices estão conectados entre si por caminhos e que ao mesmo tempo não estão conectados a mais nenhum outro vértice no supergrafo.

No grafo apresentado na figura 6 acima, o algoritmo é iniciado escolhendo-se inicialmente um dos vértices do mesmo (no caso de um grafo desconexo, esse vértice inicial pode ser escolhido em qualquer uma das suas componentes conexas).

Na 1º componente conexa do grafo acima existem ciclos.

No exemplo, se irá escolher inicialmente o vértice com rótulo 1.

Este vértice irá para a LVPC.

Uma vez que o vértice 1 tenha entrado na lista, se atribuirá o número zero como primeira coordenada do vetor a ele correspondente. Como este vértice é o primeiro vértice a entrar nesta lista, ele será o vértice de referência para todos os demais vértices diretamente ligados a ele.

A seguir devem-se identificar todos os vértices vizinhos do vértice 1 (isto é, pertencentes a 1-vizinhança) e se escolher um deles para entrar na lista. Verifica-se que estes vizinhos são os vértices 2, 3 e 4. Escolhe-se arbitrariamente o vértice 2 para ser codificado. Quando este vértice entra na lista ele repete a estrutura (isto é, a sequência de números que representam as componentes do vetor) do seu vértice de referência (vértice 1) até o

momento. Isso corresponde a repetir a única componente (de valor zero) do vetor representando o vértice um.

A figura 7 abaixo ilustra a identificação da 1-vizinhança, nesta etapa do algoritmo.

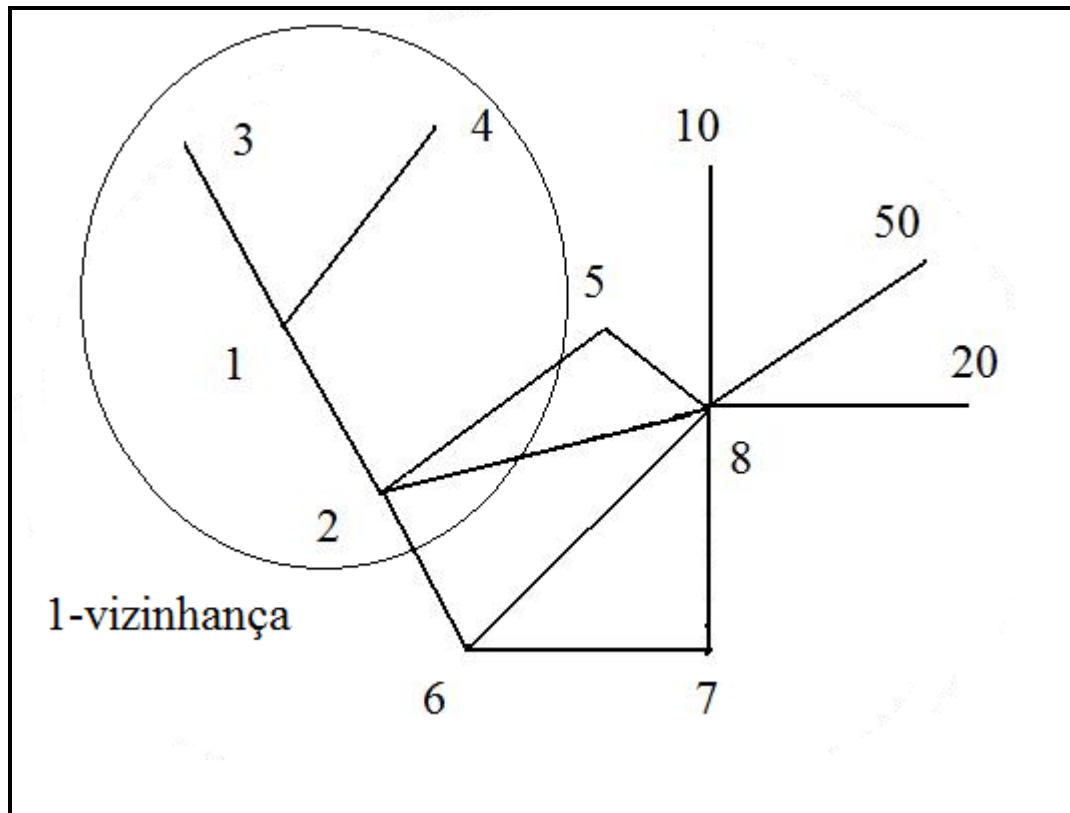


Figura 7: Identificação da 1-vizinhança no 1º componente conexo do grafo.

A cada vértice que entra na lista de vértices em processo de codificação, o vetor que o representa deve, em sua estrutura, repetir as componentes até a iteração vigente do vetor representando o vértice utilizado no momento para sua codificação. Uma vez repetida esta estrutura, deve-se acrescentar mais uma coordenada neste vetor, representando este novo vértice na lista, com valor unitário. Como este vetor terá uma dimensão a mais dos demais já incluídos anteriormente nesta lista, deve-se aumentar a dimensão deles todos, com o valor de zero.

Assim, após a inclusão dos vértices 1, 2, 3 e 4 do grafo na lista (vértices estes que entraram na lista na ordem mencionada), se obterá a seguinte estrutura dos vetores os representando:

LVPC

Vértice 1: [0, 0, 0, 0

Vértice 2: [0, 1, 0, 0

Vértice 3: [0, 0, 1, 0

Vértice 4: [0, 0, 0, 1

Figura 8: LVPC após inclusão dos vértices 2, 3, e 4.

Após a inclusão destes vértices na LVPC não há mais 1-vizinhos a serem incluídos na lista. Deve-se então, para cada um desses 1-vizinhos, se identificar e codificar todos os vértices adjacentes a eles. Neste sentido, identifica-se então que o vértice 2 (que foi o primeiro dos 1-vizinhos a entrar na LVPC) é o único que possui vértices adjacentes a ele, e que estes vértices são os rotulados pelos números 5, 8 e 6.

A figura 9 abaixo expressa esta etapa do algoritmo.

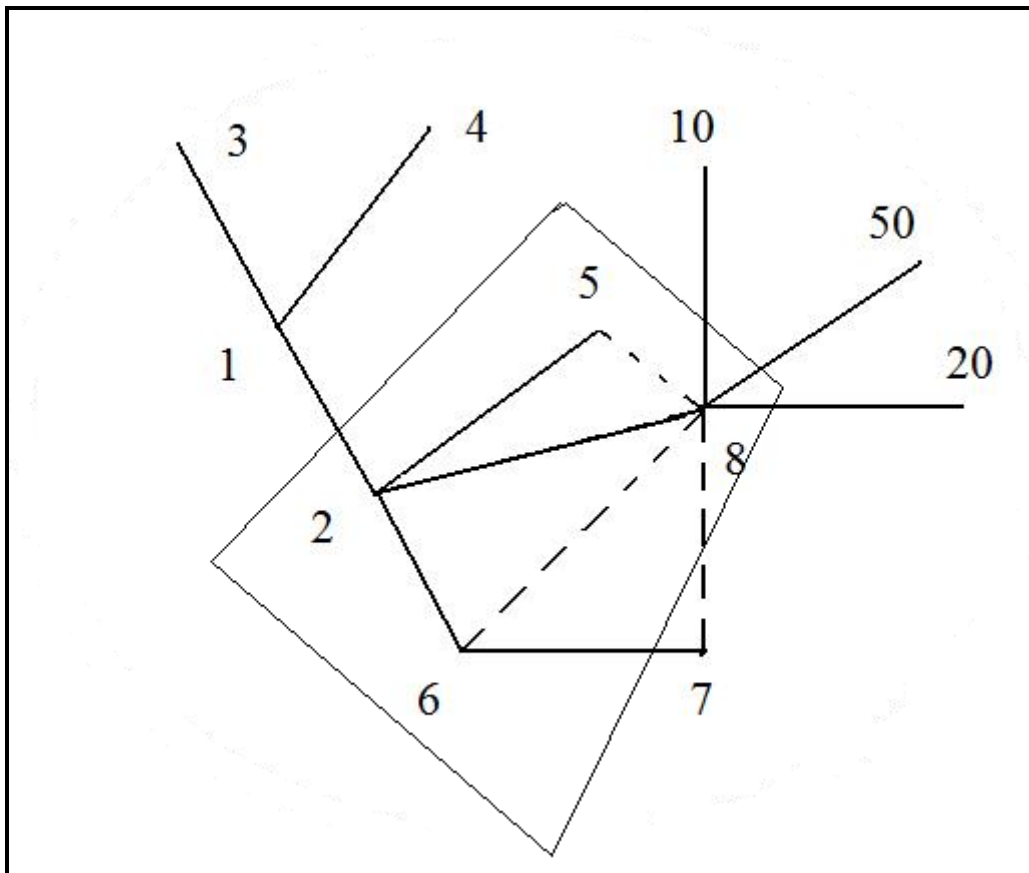


Figura 9: Identificação dos vértices adjacentes ao 1-vizinho 2 no 1º componente conexo do grafo.

Até então o vértice dois está codificado como [0, 1, 0, 0 e os seus vértices adjacentes devem repetir esta estrutura. O vértice dois passa então a ser o vértice de referência para codificação destes vértices. Assim, se determinará que o primeiro vértice adjacente ao

vértice dois a entrar na lista é o vértice cinco. Como o vetor representando este vértice deverá repetir a estrutura de 2 até então, acrescido de uma coordenada de valor unitário, sua estrutura, nesta etapa é apresentada abaixo, conjuntamente com a estrutura dos demais vetores nesta interação:

LVPC
Vértice 1: [0, 0, 0, 0, 0
Vértice 2: [0, 1, 0, 0, 0
Vértice 3: [0, 0, 1, 0, 0
Vértice 4: [0, 0, 0, 1, 0
Vértice 5: [0, 1, 0, 0, 1

Figura 10: LVPC após inclusão do vértice 5.

Como se pode verificar, ao se incluir um vetor novo na lista, deve-se aumentar a dimensão dos demais vetores nesta lista acrescentando-se mais uma coordenada, sempre com valor de zero.

Seguindo-se a lógica do algoritmo, os vértices 8 e 6 são incluídos (nesta ordem) na lista de codificação. Assim, após estas duas etapas tem-se a lista atualizada da seguinte forma:

LVPC
Vértice 1:[0, 0, 0, 0, 0, 0, 0
Vértice 2: [0, 1, 0, 0, 0, 0, 0
Vértice 3: [0, 0, 1, 0, 0, 0, 0
Vértice 4: [0, 0, 0, 1, 0, 0, 0
Vértice 5: [0, 1, 0, 0, 1, 0, 0
Vértice 8: [0, 1, 0, 0, 0, 1, 0
Vértice 6: [0, 1, 0, 0, 0, 0, 1

Figura 11: LVPC após inclusão dos vértices 6 e 8.

Como não há mais vizinhos do vértice 2 para entrar na lista, deve-se partir para a codificação dos vértices diretamente ligados aos vizinhos dos outros vértices que já entraram na lista (ou seja, vértices 3 e 4). Como não há vizinhos destes vértices, deve-se prosseguir na codificação dos vizinhos dos demais vértices.

Neste sentido, o próximo vértice para se verificar se possui vizinho ou não é o vértice 8. Verifica-se que este vértice possui como vizinhos os vértices 10, 50 e 20. Estes vértices entrarão na lista (um de cada vez) aproveitando a estrutura do vetor representando o vértice

8 até esta etapa e repetindo-se as etapas anteriormente explicadas. Deste modo, este último vértice passará a ser utilizado como referência na codificação destes três vértices.

Esta etapa é ilustrada na figura 12 abaixo:

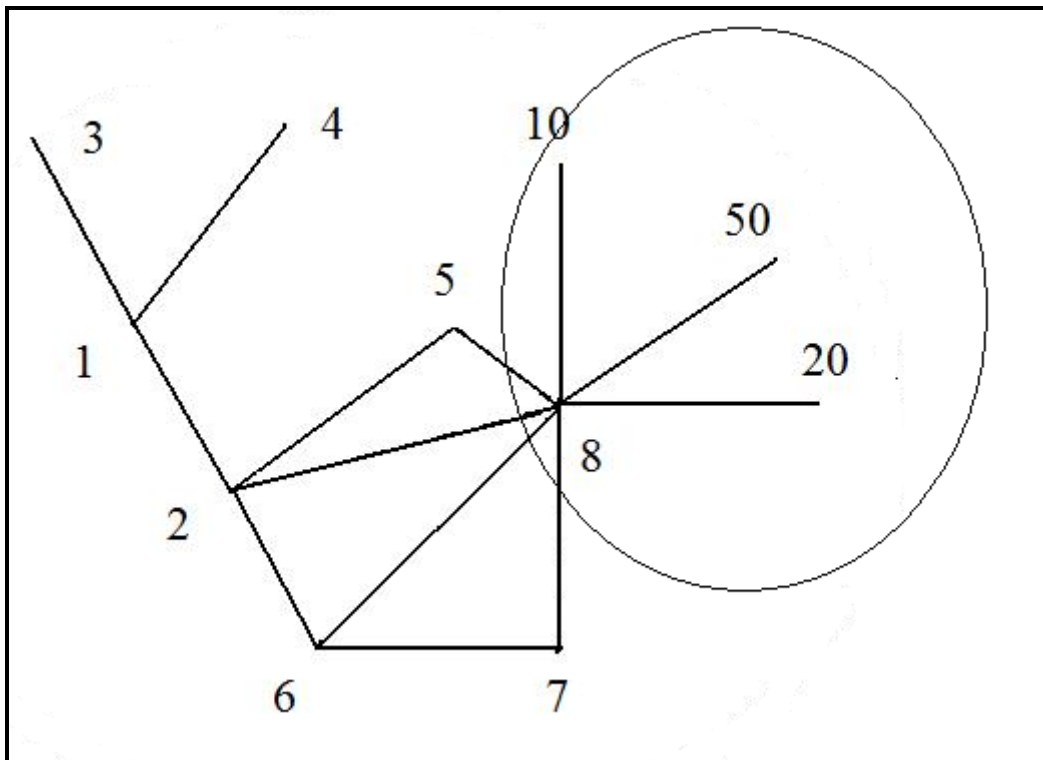


Figura 12: Identificação da 8-vizinhança (vértices vizinhos ao vértice 8) no 1º componente conexo do grafo.

Assim, após a entrada destes vértices na lista de codificação tem-se a seguinte configuração:

LVPC	
Vértice 1:	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 2:	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 3:	[0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 4:	[0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
Vértice 5:	[0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0]
Vértice 8:	[0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0]
Vértice 6:	[0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0]
Vértice 10:	[0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0]
Vértice 50:	[0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0]
Vértice 20:	[0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0]

Figura 13: LVPC após inclusão dos vértices 10, 20 e 50.

Resta checar se existem vizinhos do vértice 6. Este vértice possui como vizinhos os vértices 8 (já incluído na lista de codificação) e o vértice 7.

Deste modo deve-se incluir na lista de codificação o vértice 7 (tendo-se como referência para a codificação do mesmo o vértice 6). Com esta inclusão tem-se a seguinte configuração:

LVPC	
Vértice 1:	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 2:	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 3:	[0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 4:	[0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 5:	[0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
Vértice 8:	[0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]
Vértice 6:	[0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
Vértice 10:	[0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0]
Vértice 50:	[0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0]
Vértice 20:	[0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0]
Vértice 7:	[0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1]

Figura 14: LVPC após a inclusão de todos os vértices do 1º componente conexo do grafo.

Após esta etapa verifica-se que não há mais vizinhos dos vértices recém-incluídos na lista (vértices 10, 50, 20 e 7). Esta situação indica duas possibilidades: ou não há mais vértices para serem codificados no vértice inteiro, ou existem outras componentes conexas neste grafo (que é o caso do exemplo). Para tanto se deve verificar (por exemplo, em uma lista de adjacência do grafo) se existem mais vértices neste grafo que ainda não foram incluídos na lista de codificação. Caso positivo, este grafo possui mais de uma componente conexa e o algoritmo prosseguirá.

O grafo do exemplo possui duas componentes conexas. Deve-se guardar a codificação até o momento dos vértices na primeira componente conexa e se partir para a codificação dos vértices da segunda componente.

Nesta segunda componente conexa, verifica-se que existem 4 vértices: os vértices 200, 201, 100 e 80. Deve-se escolher qualquer um deles para ser o vértice de referência na codificação dos demais. Para tanto, no exemplo, escolhe-se o vértice 200, que entrará na lista de codificação.

Deve-se observar aqui que, antes da inclusão do vértice 200, já havia na lista 11 vetores (cada um deles com 11 coordenadas). A figura 15 abaixo ilustra a identificação da 20-vizinhança no segundo componente conexo.

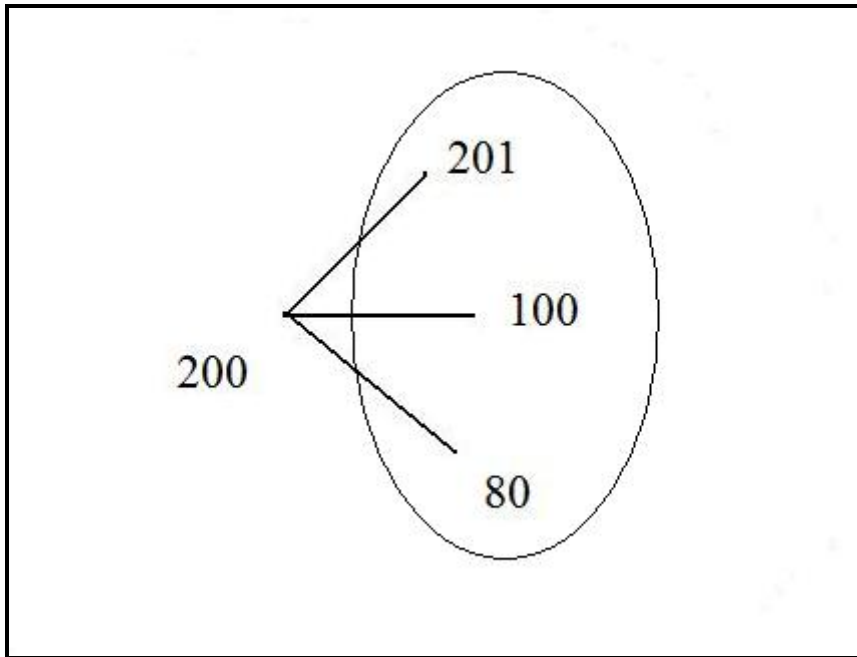


Figura 15: Identificação da 20-vizinhança na 2ª componente conexa.

Deste modo, quando o vértice 200 entrar na lista, será a décimo segundo a ser codificado e deverá ter, como estrutura, 12 componentes (todas com valor zero). Como anteriormente, deve-se adicionar uma coordenada unitária representando este vértice e uma coordenada nula em todos os demais vetores na lista para que tenham a mesma dimensão do vetor representando o vértice 200. Todas as etapas descritas anteriormente se repetem, com a inclusão dos vértices vizinhos de 200. Com a inclusão destes, não há mais vértices a serem codificados no grafo, e os vetores os representando até este momento estão estruturados da seguinte forma:

A figura abaixo retrata a LVPC neste momento do processo de codificação.

levando-se em consideração o número total de vértices no grafo e o número de componentes conexas no mesmo e deve variar para cada uma dessas componentes.

Assim, como se tem 15 vértices no grafo (11 no primeiro componente conexo e 4 no segundo) escolheu-se o valor de 500 para a primeira componente conexa e o valor de 1000 para a segunda.

No mesmo exemplo, como o grafo considerado possui 2 componentes conexas, serão incluídos mais duas coordenadas adicionais em cada vetor na lista.

Assim, os vetores correspondentes a primeira componente conexa do grafo receberão as coordenadas adicionais [500, 0] no início da sequência dos componentes dos mesmos.

Similarmente, os vetores da segunda componente receberão o par [0, 1000].

Com esta etapa conclui-se a codificação dos vértices no grafo. Desta forma os vetores representando estes vértices são assim formados:

LVPC	
Componente conexo 1:	
Vértice 1:	[500, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 2:	[500, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 3:	[500, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 4:	[500, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 5:	[500, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 8:	[500, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 6:	[500, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
Vértice 10:	[500, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0]
Vértice 50:	[500, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0]
Vértice 20:	[500, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0]
Vértice 7:	[500, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0]
Componente conexo 2:	
Vértice 200:	[0, 1000, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 201:	[0, 1000, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0]
Vértice 100:	[0, 1000, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]
Vértice 80:	[0, 1000, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]

Figura 17: LVPC após final do processo de codificação.

Pelos motivos apresentados anteriormente a codificação proposta acima dos vértices de um grafo é perfeitamente aplicável a grafos que sejam árvores ou florestas, pois, como estes

grafos não possuem ciclos, o quadrado da distância euclidiana entre dois vetores que representem vértices em uma mesma componente conexa representa exatamente a menor distância entre estes vértices (com arestas com peso 1 ligando estes vértices).

Com esta mesma codificação, a distância entre dois vértices pertencentes a diferentes componentes conexas, é muito maior do que as distâncias entre vértices no mesmo componente (o que é fundamental em problemas de clusterização).

Finalmente, vale notar que, com a metodologia de codificação proposta, aliada a algoritmos de clusterização que façam uso dos vetores assim construídos, se transformou o problema de clusterização em grafos em um problema de clusterização de vetores em um espaço de alta dimensão.

A mesma metodologia de codificação pode ser generalizada para n componentes conexas.

Exemplo 2: O exemplo a seguir apresentado ilustra a utilização do algoritmo de codificação em uma árvore.

A figura 18 abaixo apresenta uma árvore composta de cinco vértices.

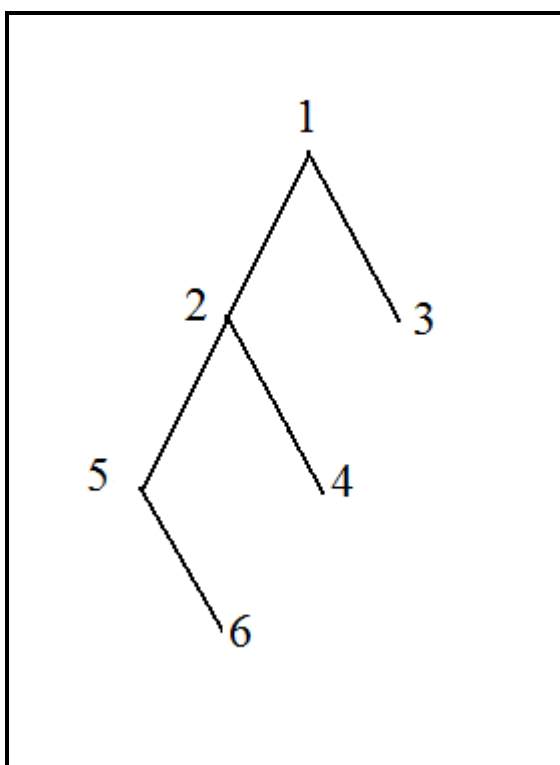


Figura18: Árvore composta por cinco vértices.

Aplicando-se o algoritmo de codificação, iniciando-se o processo de codificação pelo vértice 1 e supondo que a ordem de entrada dos vértices na LVPC seja 1, 3, 2, 4, 5 e 6 se obterá os seguintes vetores:

LVPC
Vértice 1: [0, 0, 0, 0, 0, 0]
Vértice 3: [0, 1, 0, 0, 0, 0]
Vértice 2: [0, 0, 1, 0, 0, 0]
Vértice 4: [0, 0, 1, 1, 0, 0]
Vértice 5: [0, 0, 1, 0, 1, 0]
Vértice 6: [0, 0, 1, 0, 1, 1]

Figura 19: LVPC após o final da codificação do exemplo 2.

Computando-se as distâncias euclidianas e o quadrado dessas distâncias é imediato verificar que esses valores são os apresentados nas matrizes abaixo.

Tabela 1: Distâncias euclidianas entre os vetores representativos dos vértices do grafo apresentado.

	1	2	3	4	5	6
1	0,00	1,00	1,00	1,41	1,41	1,73
2	1,00	0,00	1,41	1,00	1,00	1,41
3	1,00	1,41	0,00	1,73	1,73	2,00
4	1,41	1,00	1,73	0,00	1,41	1,73
5	1,41	1,00	1,73	1,41	0,00	1,00
6	1,73	1,41	2,00	1,73	1,00	0,00

Tabela 2: Quadrado das distâncias euclidianas entre os vértices da árvore considerada.

	1	2	3	4	5	6
1	0,00	1,00	1,00	2,00	2,00	3,00
2	1,00	0,00	2,00	1,00	1,00	2,00
3	1,00	2,00	0,00	3,00	3,00	4,00
4	2,00	1,00	3,00	0,00	2,00	3,00
5	2,00	1,00	3,00	2,00	0,00	1,00
6	3,00	2,00	4,00	3,00	1,00	0,00

Deste modo, pode-se facilmente verificar que, na ausência de ciclos, o quadrado das distâncias euclidianas entre os vértices do grafo, computadas a partir dos vetores gerados pelo algoritmo, fornece o número de passos necessários, para, partindo-se de um vértice qualquer na árvore, chegar ao outro.

2.2.1.4 Outras características da metodologia de codificação proposta

- A metodologia proposta possui etapas que podem ser separadas, o que pode abrir a possibilidade para o uso de computação paralela, ou *multithreading*, desde que se tenha um conhecimento prévio do número de componentes conexas do grafo (e que o mesmo seja

suficientemente grande para justificar a implementação de *multithreading* ou outras técnicas similares).

Isto ocorre porque o processo de codificação em cada componente conexa é em boa parte independente do processo nas demais componentes (devendo somente levar em consideração o número de coordenadas nos vetores na lista, em cada instante).

Por exemplo, quando o vértice 200 entrou na lista de codificação somente se repetiu o número de coordenadas –todas nulas – dos vetores até então na lista e se adicionou um zero como nova coordenada do vetor correspondente a este vértice; o que corresponde a se iniciar o processo de codificação de novo, nesta componente conexa.

Isto significa que se pode simplesmente iniciar a codificação nesta componente independentemente das codificações nas demais componentes conexas e depois se incluir, no início da lista de coordenadas de cada vetor nesta componente conexa, o número de casas (coordenadas)– todas iguais a zero – correspondente ao número de vetores na lista.

- Pode incluir uso intensivo de memória RAM: dependendo das dimensões do grafo em que se aplica a codificação proposta, os vetores resultantes podem implicar em uso de quantidades substanciais da memória RAM disponível para se carregar tais vetores para que possam ser utilizados em um determinado algoritmo de clusterização.

2.2.1.5 A codificação proposta e a maldição da dimensionalidade

A codificação proposta implica que, um algoritmo que empregue diretamente os vetores assim construídos (representando os vértices de um grafo) terá que utilizar diretamente vetores de dimensão elevada. Se o grafo de interesse for muito grande, problemas derivados da "maldição da dimensionalidade" podem ser verificados.

Para tanto, faz-se necessário primeiramente definir o que é a maldição da dimensionalidade e quais suas consequências em aplicações que façam uso dessa codificação.

2.2.1.6 A maldição da dimensionalidade

A maldição da dimensionalidade se refere a vários fenômenos que se verificam na análise e organização de dados em espaços de alta dimensão (frequentemente em espaços de dimensões da ordem de milhar) e que não ocorrem em espaços de baixa dimensão.

Esses fenômenos se verificam em uma série de aplicações; como por exemplo, análise numérica, amostragem, combinatória, aprendizagem de máquinas, *data mining* etc.

O que se verifica, como fenômeno, em todas estas aplicações é que quando a dimensionalidade aumenta os dados se tornam esparsos. A esparsidade é problemática em

métodos que requerem significância estatística (nestes métodos, para se obter um resultado confiável estatístico, a quantidade de dados requerida para suportar este resultado frequentemente cresce exponencialmente com o aumento da dimensionalidade).

Em problemas envolvendo representações matriciais, a esparcidade significa que as matrizes associadas contêm uma grande quantidade de elementos nulos (o que significa que somente um pequeno número de ramos da rede são conectados a cada nó). Portanto, processando e armazenando apenas os elementos não nulos pode-se melhorar consideravelmente a eficiência computacional (COLEMAR, 2013). Este é o caso do algoritmo de codificação dos vértices de um grafo proposto; pois os vetores criados pelo mesmo são estruturas muito esparsas (compostos em sua maioria de componentes nulos).

2.2.1.7 Outros problemas identificados com a maldição da dimensionalidade

Além dos problemas levantados acima, outros fenômenos foram identificados com o aumento da dimensionalidade dos dados. A seguir se apresentará de forma sucinta alguns problemas identificados em diversos tipos de aplicações.

- Aprendizagem de máquinas (*Machine Learning*):

Em problemas de aprendizagem de máquinas há a preocupação de se aprender um “estado da natureza” de uma amostra de dados provindos de um espaço de *features* de alta dimensão, onde cada *feature* assume um determinado número de valores.

Conforme a dimensionalidade aumenta desse espaço, uma grande quantidade de dados para treinamento é requerida para se assegurar que na amostra existe cada combinação de valor possível. Mantendo-se fixo o número de elementos na amostra utilizada para treinamento da máquina de aprendizado, o seu poder preditivo de reduz à medida que a dimensionalidade dos dados aumenta (fenômeno de Hughes).

Nos problemas de clusterização de palavras em *thesauri* não faz sentido, a princípio, se preocupar com o poder preditivo das máquinas de aprendizado encarregadas destas tarefas uma vez que não existe uma única resposta possível (ou uma única configuração de *clusters* aceitável por todos).

No caso dos algoritmos utilizados neste trabalho os *clusters* foram determinados por algoritmos que assumem claramente esta premissa (o que, no caso de redes neurais de Kohonen é traduzido por meio de um treinamento não supervisionado, ou seja, um treinamento onde se assume que não há uma resposta única correta (alvo) a ser apresentada à rede durante o seu treinamento).

- Estatística Bayesiana:

Nos problemas envolvendo inferência Bayesiana a determinação da distribuição a posteriori também enfrenta problemas decorrentes de alta dimensionalidade dos dados. Para a determinação desta distribuição é necessário frequentemente se efetuar uma integração de uma função de alta dimensão (podendo chegar a dimensões da ordem de milhar ou de dezena de milhar) para a qual não se conhece nenhuma técnica de integração conhecida.

Adicionalmente, os métodos numéricos de integração não produzem bons resultados em espaços de dimensão tão elevada. Este problema é enfrentado por meio da determinação da distribuição a posteriori através de métodos MCMC (*Markov Chain Monte Carlo*). No entanto, estes métodos baseados em simulação convergem lentamente, o que representa um problema da alta dimensionalidade dos dados.

- Funções distância:

Quando uma medida de distância (como por exemplo, a distância Euclidiana) é utilizada com dados de alta dimensão existe pouca diferença entre diferentes pares de amostras. Isto representa um problema grave para algoritmos que necessitam auferir a distância entre elementos de um espaço de dimensão muito elevada.

Para fins de ilustração deste fato, basta verificar como a proporção do hiper-volume de uma hiper-esfera com o hiper-volume de um hiper-cubo varia conforme a dimensão do espaço vai aumentando.

Neste sentido, sejam:

r: o raio da hiper-esfera,

d: a dimensão do espaço

O hipercubo possui lados de comprimento $2r$, em um espaço de dimensão d .

O hiper-volume da esfera é dado por: $\frac{2r^d \pi^{\frac{d}{2}}}{d \Gamma\left(\frac{d}{2}\right)}$.

O hiper-volume do cubo é dado por: $(2r)^d$.

O limite da razão entre os dois hiper-volumes (razão definida com o volume da hiper-esfera como numerador dessa razão), conforme a dimensão d do espaço aumenta é zero; isto

$$\text{é: } \frac{\pi^{\frac{d}{2}}}{d 2^{d-1} \Gamma\left(\frac{d}{2}\right)} \rightarrow 0 \text{ conforme } d \rightarrow \infty$$

Isto significa que o volume da hiper-esfera vai representando um volume cada vez mais insignificante em relação ao volume do hiper-cubo. Logo, um espaço de dimensão

muito elevada pode ser encarado como consistindo quase que inteiramente de “cantos” ou “quinas” (*corners*) de um hiper-cubo, com quase nada no “meio”.

Neste espaço, as distâncias máximas e mínimas se tornam indiscerníveis e convergem à zero na medida em que a dimensão aumenta; isto é: $\lim_{d \rightarrow \infty} \frac{\text{dist}_{\max} - \text{dist}_{\min}}{\text{dist}_{\min}} \rightarrow 0$

Isto significa que as funções de distância perdem sua utilidade conforme a dimensão d aumenta muito.

2.2.1.8 Consequências da maldição da dimensionalidade para algoritmos que utilizem a codificação proposta dos vértices de um grafo acíclico

Dos fenômenos associados à maldição da dimensionalidade levantados anteriormente os relevantes para a utilização da metodologia de codificação dos vértices de um grafo de porte muito grande são os seguintes:

- Esparsidade dos vetores construídos: os vetores representando os vértices de um grafo acíclico muito grande (contendo milhares ou dezenas de milhares de vértices) serão necessariamente muito esparços. Isto significa que procedimentos computacionais específicos terão que ser utilizados para lidar com a esparsidade dessas estruturas. Em particular devem-se adotar procedimentos que só realizem operações com os elementos não nulos desses vetores.

A esse respeito diversas linguagens de programação (como por exemplo, a linguagem MATLAB e diversas bibliotecas de manipulação de matrizes em Java) oferecem recursos para a realização de operações somente com elementos não nulos de matrizes e vetores. Mais ainda, estas linguagens armazenam também na memória RAM do computador somente as parcelas não nulas dessas estruturas; melhorando a desempenho de algoritmos que usem estruturas tão esparsas.

- O problema da alta dimensionalidade de dados com a utilização de funções distância: este problema é um problema sério para algoritmos que utilizem diretamente os vetores construídos segundo o esquema proposto, em grafos de porte muito grande, conjuntamente com uma função distância. Neste caso não há solução aparente para os problemas associados aos cálculos de distâncias entre elementos de uma amostra de um espaço de dimensão muito elevada.

As implicações deste fato para algoritmos de clusterização de dados que empregam diretamente medidas de distância em dados de alta dimensão são imediatas, na medida em que a utilidade dessas medidas cai conforme a dimensão do espaço aumenta. Assim, utilizar

algoritmos de clusterização como o *K-means* com vetores de altíssima dimensão não produzirá resultados úteis.

Por outro lado, este problema pode ser contornado em algoritmos de clusterização que promovam uma redução na dimensão dos vetores assim construídos.

2.2.1.9 Aplicação da metodologia proposta em grafos com ciclos

Este trabalho testou o algoritmo proposto em grafos com uma grande variedade de configurações de ciclos, inclusive aquelas onde existam ciclos compartilhando vértices e arestas. Nestas aplicações verificou-se que as distâncias computadas entre pares de vetores representando vértices pertencentes a ciclos com estas características aproximam bem as distâncias mínimas entre estes vértices.

O exemplo a seguir ilustrará a aplicação desse algoritmo em um tipo de configuração de ciclos testada, de forma a ilustrar os bons resultados obtidos com o mesmo.

2.2.2.0 Exemplo de aplicação do algoritmo de codificação proposto em um grafo contendo ciclos que compartilham vértices

Esta sessão irá apresentar a aplicação do algoritmo proposto em um grafo contendo ciclos dentro de ciclos.

A figura abaixo apresenta como os vértices e ciclos do grafo estão dispostos.

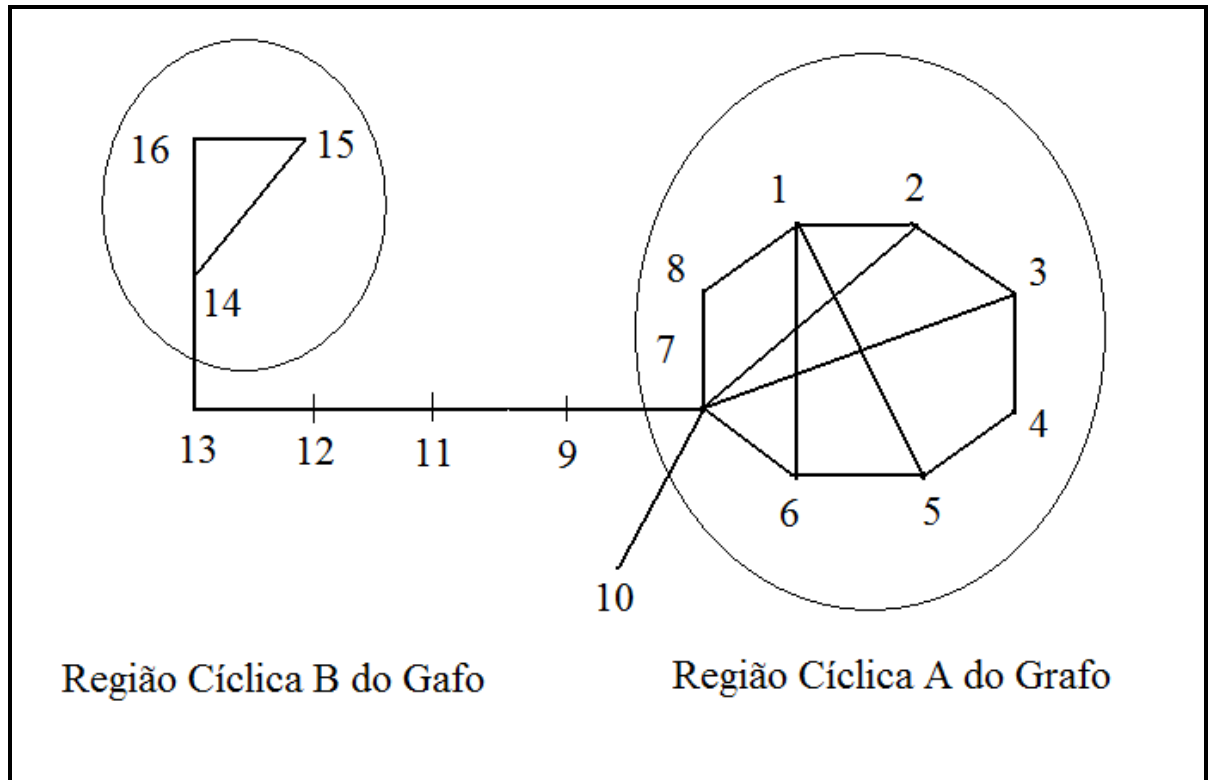


Figura 20: Configuração dos ciclos do grafo a ser utilizado no algoritmo de codificação

A figura 20 acima apresenta um grafo contendo duas regiões cíclicas. A região A contém diversos sub-ciclos e a região B é composta por um único ciclo.

Procedendo-se à codificação destes vértices com o algoritmo proposto, iniciando-se o processo no vértice 4 e admitindo-se que a ordem de entrada dos vértices na LVPC seja 4, 3, 5, 7, 2, 6, 1, 9, 10, 8, 11, 12, 13, 14, 15 e 16, pode-se verificar que ao final do processo de codificação, os vértices serão representados pelos seguintes vetores:

LVPC	
Vértice 04:	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 03:	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 05:	[0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 07:	[0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 02:	[0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 06:	[0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 01:	[0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 09:	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
Vértice 10:	[0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
Vértice 08:	[0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]
Vértice 11:	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0]
Vértice 12:	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0]
Vértice 13:	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0]
Vértice 14:	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0]
Vértice 15:	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0]
Vértice 16:	[0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1]

Figura 21: Vetores representando os 16 vértices do grafo utilizado como exemplo, ao final do processo de codificação.

Como os vértices considerados se encontram todos na mesma (e única) componente conexa, não houve necessidade de se incluir coordenadas adicionais nos vetores.

As tabelas 3 a 5 a seguir apresentam as matrizes de distâncias mínimas entre os pares de vértices do grafo apresentado, as distâncias euclidianas entre seus vetores representantes e o quadrado dessas distâncias euclidianas.

Tabela 3: Distâncias euclidianas entre os vetores representativos dos vértices do grafo apresentado.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0,00	2,00	1,73	1,41	1,00	1,41	2,00	2,24	1,73	2,24	2,00	2,65	2,83	3,00	3,16	3,16
2	2,00	0,00	1,00	1,41	1,73	2,00	1,41	1,73	1,73	1,73	2,00	2,24	2,45	2,65	2,83	2,83
3	1,73	1,00	0,00	1,00	1,41	1,73	1,00	1,41	1,41	1,41	1,73	2,00	2,24	2,45	2,65	2,65
4	1,41	1,41	1,00	0,00	1,00	1,41	1,41	1,73	1,73	1,73	2,00	2,24	2,45	2,65	2,83	2,83
5	1,00	1,73	1,41	1,00	0,00	1,00	1,73	2,00	2,00	2,00	2,24	2,45	2,65	2,83	3,00	3,00
6	1,41	2,00	1,73	1,41	1,00	0,00	1,73	2,24	2,24	2,24	2,45	2,65	2,83	3,00	3,16	3,16
7	2,00	1,41	1,00	1,41	1,73	1,73	0,00	1,00	1,00	1,00	1,41	1,73	2,00	2,24	2,45	2,45
8	2,24	1,73	1,41	1,73	2,00	2,24	1,00	0,00	1,41	1,41	1,73	2,00	2,24	2,45	2,65	2,65
9	1,73	1,73	1,41	1,73	2,00	2,24	1,00	1,41	0,00	1,41	1,00	1,41	1,73	2,00	2,24	2,24
10	2,24	1,73	1,41	1,73	2,00	2,24	1,00	1,41	1,41	0,00	1,73	2,00	2,24	2,45	2,65	2,65
11	2,45	2,00	1,73	2,00	2,24	2,45	1,41	1,73	1,00	1,73	0,00	1,00	1,41	1,73	2,00	2,00
12	2,65	2,24	2,00	2,24	2,45	2,65	1,73	2,00	1,41	2,00	1,00	0,00	1,00	1,41	1,73	1,73
13	2,83	2,45	2,24	2,45	2,65	2,83	2,00	2,24	1,73	2,24	1,41	1,00	0,00	1,00	1,41	1,41
14	3,00	2,65	2,45	2,65	2,83	3,00	2,24	2,45	2,00	2,45	1,73	1,41	1,00	0,00	1,00	1,00
15	3,16	2,83	2,65	2,83	3,00	3,16	2,45	2,65	2,24	2,65	2,00	1,73	1,41	1,00	0,00	1,00
16	3,16	2,83	2,65	2,83	3,00	3,16	2,45	2,65	2,24	2,65	2,00	1,73	1,41	1,00	1,00	0,00

Tabela 4: Distâncias mínimas (em termos de números de passos para se chegar a um vértice qualquer do grafo, partindo-se de outro).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	1	2	2	1	1	2	1	3	3	4	5	6	7	8	8
2	1	0	1	2	2	2	1	2	2	2	3	4	5	6	7	7
3	2	1	0	1	2	2	1	2	2	2	3	4	5	6	7	7
4	2	2	1	0	1	2	2	3	3	3	4	5	6	7	8	8
5	1	2	2	1	0	1	2	2	3	3	4	5	6	7	8	8
6	1	2	2	2	1	0	1	2	2	2	3	4	5	6	7	7
7	3	1	1	2	2	1	0	1	1	1	2	3	4	5	6	6
8	1	2	2	3	2	2	1	0	2	2	3	4	5	6	7	7
9	3	2	2	3	3	2	1	2	0	2	1	2	3	4	5	5
10	3	2	2	3	3	2	1	2	2	0	3	4	5	6	7	7
11	4	3	3	4	4	3	2	3	1	3	0	1	2	3	4	4
12	5	4	4	5	5	4	3	4	2	4	1	0	1	2	3	3
13	6	5	5	6	6	5	4	5	3	5	2	1	0	1	2	2
14	7	6	6	7	7	6	5	6	4	6	3	2	1	0	1	1
15	8	7	7	8	8	7	6	7	5	7	4	3	2	1	0	1
16	8	7	7	8	8	7	6	7	5	7	4	3	2	1	1	0

Tabela 5: Distâncias euclidianas ao quadrado entre os vetores representativos dos vértices do grafo apresentado.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	4	3	2	1	2	4	5	5	5	6	7	8	9	10	10
2	4	0	1	2	3	4	2	3	3	3	4	5	6	7	8	8
3	3	1	0	1	2	3	1	2	2	2	3	4	5	6	7	7
4	2	2	1	0	1	2	2	3	3	3	4	5	6	7	8	8
5	1	3	2	1	0	1	3	4	4	4	5	6	7	8	9	9
6	2	4	3	2	1	0	3	5	5	5	6	7	8	9	10	10
7	4	2	1	2	3	3	0	1	1	1	2	3	4	5	6	6
8	5	3	2	3	4	5	1	0	2	2	3	4	5	6	7	7
9	5	3	2	3	4	5	1	2	0	2	1	2	3	4	5	5
10	5	3	2	3	4	5	1	2	2	0	3	4	5	6	7	7
11	6	4	3	4	5	6	2	3	1	3	0	1	2	3	4	4
12	7	5	4	5	6	7	3	4	2	4	1	0	1	2	3	3
13	8	6	5	6	7	8	4	5	3	5	2	1	0	1	2	2
14	9	7	6	7	8	9	5	6	4	6	3	2	1	0	1	1
15	10	8	7	8	9	10	6	7	5	7	4	3	2	1	0	1
16	10	8	7	8	9	10	6	7	5	7	4	3	2	1	1	0

Pode-se verificar, pela análise das matrizes de distâncias computadas entre os pares de vetores representando os vértices do grafo usado neste exemplo que:

a) As distâncias computadas entre pares de vértices pertencentes a uma mesma região cíclica do grafo foram menores do que as distâncias computadas entre pares onde cada um dos vértices pertença a diferentes regiões cíclicas.

Tal padrão se verificou ao longo dos diferentes experimentos realizados com o algoritmo de codificação, utilizado em conjunto com redes SOM (incluindo o experimento envolvendo traços de personalidade retirados de um *thesaurus*, a ser apresentado no capítulo 7). Este padrão é desejável quando se utiliza tal algoritmo na produção de dados de *input* para um algoritmo de clusterização; uma vez que contribuirá para que o algoritmo aloque os vértices pertencentes a cada uma dessas regiões em *clusters* diferentes.

b) Nos trechos sem ciclos do grafo, o quadrado das distâncias euclidianas entre vértices pertencentes a estes trechos, representa a distância mínima entre os mesmos.

c) As distâncias atribuídas (por meio dos vetores) entre vértices de uma mesma região cíclica não foram muito diferentes das distâncias mínimas entre estes vértices, para a grande maioria dos casos.

d) Para o vértice de início do processo de codificação (vértice 4), os quadrados das distâncias entre este vértice e qualquer outro representaram exatamente as menores

distâncias, mesmo considerando-se as distâncias computadas em relação aos vértices pertencentes a ciclos.

e) As distâncias euclidianas atribuídas entre pares de vértices, onde um desses vértices pertence a uma região cíclica e o outro não, é proporcional às distâncias mínimas entre esses pares. Assim, fixando-se um vértice pertencente a uma região cíclica, e computando-se sucessivamente as distâncias euclidianas entre este vértice e vértices não pertencentes a essas regiões (em uma sequência tal, na qual os vértices não pertencentes a ciclos sejam considerados nesses cálculos por ordem crescente de distância mínima em relação ao vértice fixado), as distâncias computadas também apresentarão valores crescentes nesta sequência.

f) As distâncias euclidianas atribuídas entre pares de vértices, onde um desses vértices pertença a uma região cíclica e o outro pertença a outra região cíclica, são proporcionais às distâncias mínimas entre esses pares, na esmagadora maioria dos casos.

Estas características foram observadas nos testes da metodologia de codificação de vértices proposta, em uma grande variedade de grafos diferentes (cada um deles com diferentes configurações de regiões cíclicas); incluindo o grafo composto por 222 vértices, representando as ligações entre traços de personalidade segundo o *True Term Thesaurus Database* (razão pela qual esta metodologia foi utilizada, em conjunto com redes neurais SOM, na clusterização desse grafo; como será apresentado nos capítulos 6 e 7 desta tese).

Assim, como se pode observar no experimento acima, embora as distâncias euclidianas computadas (a partir dos vetores gerados pelo algoritmo apresentado) entre vértices de uma mesma região cíclica sejam, em diversos casos, diferentes das distâncias mínimas, o algoritmo aqui apresentado pode ser utilizado em problemas de clusterização onde se deseje agrupar em um mesmo cluster, todos (ou a maior parte) os vértices de uma mesma região cíclica.

Estes mesmos padrões foram observados em diversos experimentos contendo grafos com variadas configurações de *clusters*.

2.2.2.1 Um exemplo de código em Java aplicado ao algoritmo de codificação

Um possível código (comentado) implementando as rotinas de codificação proposta para os vértices do grafo do exemplo utilizado é apresentado no CD anexo a tese. Este código será apresentado em uma sequência de figuras para que os formatos e sessões expansíveis do código Java não sejam perdidos em texto.

2.2.2.2 Considerações finais sobre o algoritmo proposto

Com relação ao algoritmo proposto, não se encontrou indício algum de que já tenha sido desenvolvido por algum outro autor anteriormente. A esse respeito, foram realizadas pesquisas no motor de buscas na *internet*, com os títulos: “*graph clustering*” “SOM”, *graph clustering* SOM, “*graph clustering*” “kohonen”, *graph clustering* kohonen, “*graph clustering*” “*vectors*” e *graph clustering vectors*, não se encontrando nenhuma solução algorítmica como a aqui apresentada.

Adicionalmente se procedeu a buscas com os mesmos temas nos sites da *IEEE Computational Intelligence Society* (<http://cis.ieee.org/>), *International Neural Network Society* (<http://www.inns.org/neural-networks-journal>), *European Neural Network Society*, (<http://www.e-nns.org/>), e no site da *Japanese Neural Network Society* (<http://www.jnns.org/english/>), e não se encontrou nenhum desenvolvimento similar ao aqui apresentado.

Isto corrobora para o posicionamento de que esta metodologia de codificação proposta neste trabalho seja uma metodologia inédita.

2.2.2.3 Esquemas alternativos de utilização do algoritmo proposto em problemas de clusterização de grafos

A seção anterior apresentou o algoritmo de codificação sendo aplicado a todas as componentes conexas do grafo, em uma única execução do algoritmo. De fato, foi com esta configuração que o algoritmo foi aplicado, neste trabalho.

Contudo, é possível se fazer usos diferentes do mesmo algoritmo em tais problemas. Um esquema alternativo de se aplicar o algoritmo pode ser vislumbrado em problemas de clusterização de grafos desconexos, e consiste em aplicações independentes do mesmo em cada uma dessas componentes.

Deste modo, cogitando-se, por exemplo, aplicar o algoritmo em um grafo com 50 componentes conexas, se executaria 50 vezes o algoritmo, uma vez para cada componente. Ou seja, uma vez que não se pode aceitar clusters envolvendo vértices desconexos, poder-se-ia restringir o foco de determinação desses agrupamentos a cada componente do grafo e tratar cada uma delas como um problema de clusterização independente.

A vantagem deste esquema de utilização do algoritmo vem do fato de que as codificações dos vértices, feitas em cada rodada de execução do algoritmo (uma para cada componente) envolveria vetores de dimensões inferiores aos obtidos com o esquema de utilização do algoritmo inicialmente proposto.

2.2.3 Algoritmo de Markov *Clustering* (MCL)

Esta sessão irá apresentar o algoritmo MCL, baseado na simulação de processos estocásticos do tipo Passeio Aleatório no grafo a ser clusterizado.

Processos estocásticos podem ser sucintamente caracterizados como uma coleção de variáveis aleatórias, geralmente listadas (ou indexadas) ao longo do tempo. Estes processos possuem uma vasta gama de aplicabilidades, especialmente na área de processos de decisão (ARRUDA, 2010). Uma apresentação introdutória sobre estes processos pode ser encontrada em REIS (2014) e WASSERMAN (2004).

O algoritmo de *Markov Clustering* foi apresentado por DONGEN (2000) e se destina a resolver problemas de clusterização especificamente em grafos. Este algoritmo realiza uma simulação de fluxos no grafo a ser clusterizado. Este procedimento simula um Passeio Aleatório (processo estocástico Markoviano conhecido na literatura como *Random Walk*) no grafo. Também é conhecido pela sigla MCL ou ainda simplesmente por *networks*.

Trata-se de um algoritmo de clusterização não supervisionado especialmente desenvolvido para problemas de clusterização em grafos e é bastante eficiente considerando-se as alternativas disponíveis para este tipo de problema.

O site <http://micans.org/mcl/> disponibiliza um aplicativo altamente otimizado envolvendo implementações deste algoritmo (disponível somente para sistema operacional Linux).

MCL pode ser usado como uma ferramenta para se identificar grupos naturais em um grafo altamente conectado. Neste sentido, este algoritmo parte da concepção de um *cluster* como sendo um subconjunto de vértices desse grafo com maior densidade de ligações entre eles. Assim, regiões com maior concentração de ligações no grafo, tendem a serem candidatas a formarem *clusters*.

Conforme as iterações do algoritmo vão se dando, as probabilidades dos fluxos atingirem e permanecerem nas regiões mais conexas vão sendo atualizadas (através das operações de *inflation*) e vão tendendo a serem aumentadas enquanto estas correspondentes probabilidades vão tendendo a serem reduzidas nas regiões menos conectadas do grafo; revelando assim grupos naturais dentro desse mesmo grafo.

A justificativa para se realizar simulações de passeios aleatórios no grafo com vistas a se identificar *clusters* é a seguinte: se fosse aleatoriamente partir de um vértice qualquer do grafo para outro (conectado ao primeiro), selecionando-se o vértice de destino também aleatoriamente, seria mais provável permanecer em uma região com maior adensamento de

ligações do grafo do que em regiões com poucas ligações entre os vértices. Deste modo, cada vértice do grafo é encarado como um estado que o processo estocástico markoviano pode assumir.

Assim, regiões onde os fluxos tendem a se dar com mais frequência são regiões com maior concentração de ligações e, portanto, regiões onde podem ser identificados *clusters*.

O algoritmo alterna fases chamadas de inflação (*inflation*) com fases de expansão (*expansion*).

Na fase de inflação a matriz de probabilidades de transição do processo estocástico é alterada, elevando-se cada elemento por uma potência positiva. A seguir, os valores de cada coluna desta matriz são normalizados.

Na fase de expansão se computa mais um produto entre a matriz de Markov da iteração vigente com a matriz de Markov inicial (que indica as probabilidades de se alcançar um vértice do grafo, partindo-se de outro, em um único passo).

A operação de inflação é responsável pelo fortalecimento (em termos de aumento de probabilidades de transição) de regiões do grafo como candidatas a *clusters* ou enfraquecimento de outras regiões (também em termos de aumento de probabilidades de transição). Esta operação impacta também no grau de granularidade dos *clusters* finais obtidos. Deste modo, o algoritmo MCL modifica as matrizes de transição do processo markoviano (antes de computar a próxima potência desta matriz) de modo a tornar mais pronunciada a diferenciação, para este algoritmo, entre as regiões com maior adensamento de ligações e as regiões com menos ligações.

Por seu turno, a operação de expansão é responsável por permitir que o fluxo simulado chegue a diferentes regiões do grafo.

Finalmente deve-se destacar que o algoritmo exige um 1-grafo não direcionado para ser clusterizado.

De forma sucinta, o algoritmo MCL pode ser assim descrito:

ALGORITMO MCL:

- Determinação do grafo não direcionado a ser clusterizado e dos parâmetros de *inflation* (parâmetro r da equação apresentada anteriormente) e de *expansion* (parâmetro e).
- Determine a matriz de adjacência do grafo a ser clusterizado.
- Inclua *auto-loops* nos vértices do grafo (somando a matriz de adjacência com a matriz do grafo com a matriz identidade, de mesma dimensão). Esta etapa visa corrigir problemas na convergência do algoritmo.

- Normalize a matriz para definição de uma matriz de probabilidades de transição do processo estocástico.

- Alterne as fases de expansão e inflação (utilizando os coeficientes definidos anteriormente) até que o processo estocástico chegue ao seu “*steady state*” (e o algoritmo atinja a convergência).

- Interprete a matriz resultado para se descobrir os *clusters* formados.

Alternativamente, existe outro critério de parada para este algoritmo, baseado na energia residual da matriz atualizada pelo algoritmo em cada iteração. A energia residual, neste sentido, é definida como a diferença da soma dos elementos dessa matriz, ao longo de suas linhas e a soma do quadrado desses elementos. Assim, determinando-se um valor máximo para esta energia residual, pode-se criar tal critério de parada.

Formalmente, o operador de *inflation* do algoritmo é definido da seguinte forma. Seja uma matriz $M \in \mathbb{R}^{k \times l}$, $M \geq 0$ e um número real não negativo r . A matriz resultante do reescalonamento de cada uma das colunas de M com coeficiente de potência r é denotada por $\Gamma_r M$ e Γ_r é chamado de operador de *inflation*, com coeficiente de potência r .

O operador $\Gamma_r : \mathbb{R}^{k \times l} \rightarrow \mathbb{R}^{k \times l}$ é definido por $(\Gamma_r M)_{pq} = (M_{pq})^r / \sum_{t=1}^k (M_{tq})^r$; donde se percebe que o operador divide cada elemento M_{pq} da matriz M (elevado a potência r) e divide esse valor pelo somatório dos elementos em sua respectiva coluna, também elevado a essa potência.

2.2.3.1 Exemplo ilustrativo do funcionamento do algoritmo MCL

Um exemplo claro, simples e sucinto da mecânica de funcionamento do algoritmo é apresentado por MACROPOL (2009).

Tabela 6: Matriz de adjacência de grafo usado como exemplo do algoritmo MCL

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1				1	1			1		
2	1	1	1		1							
3		1	1	1	1							
4			1	1				1	1		1	
5		1	1		1		1	1				
6	1					1				1		
7	1				1		1			1		
8				1	1			1	1		1	
9				1				1	1		1	1
10	1					1	1			1		
11				1				1	1		1	1
12									1		1	1

Procedendo-se à adição da matriz identidade e à normalização da matriz resultante (dividindo-se cada elemento da matriz pela soma dos elementos de sua respectiva coluna), obtêm:

Tabela 7: Matriz inicial (M) utilizada pelo MCL no exemplo utilizado:

0,20	0,25			0,33	0,25		0,25					
0,20	0,25	0,25	0,20									
	0,25	0,25	0,20	0,20								
		0,25	0,20			0,20	0,20	0,20				
	0,25	0,25	0,20	0,25	0,20							
0,20				0,33				0,25				
0,20			0,20	0,25				0,25				
		0,20	0,20		0,20	0,20	0,20		0,20			
		0,20			0,20	0,20	0,20	0,20		0,20	0,33	
0,20				0,33	0,25			0,25				
		0,20				0,20	0,20		0,20	0,20	0,33	
							0,20	0,20	0,20	0,20	0,33	

Fonte: MACROPOL, 2009, p.30.

Aplicando-se o operador de inflation com coeficiente de potência 2, obtêm o resultado abaixo:

Tabela 8: Matriz do algoritmo MCL após operação de inflação com parâmetro 2 ($\Gamma_2 M^2$).

0,380	0,087	0,027		0,077	0,295	0,201			0,320		
0,047	0,347	0,210	0,017	0,150	0,019	0,066	0,012		0,012		
0,014	0,210	0,347	0,056	0,150		0,016	0,046	0,009		0,009	
		0,027	0,087	0,302	0,062		0,184	0,143		0,143	0,083
0,058	0,210	0,210	0,056	0,406		0,083	0,046	0,009	0,019	0,009	
0,142	0,017				0,295	0,083				0,184	
0,113	0,069	0,017		0,062	0,097	0,333	0,012		0,147		
		0,017	0,069	0,175	0,049	0,016	0,287	0,143		0,143	0,083
			0,017	0,175	0,012		0,184	0,288		0,288	0,278
0,246	0,017			0,019	0,295	0,201				0,320	
			0,017	0,175	0,012		0,184	0,288		0,288	0,278
				0,044			0,046	0,120		0,120	0,278

Fonte: MACROPOL, 2009, p.30.

Procedendo-se a etapas alternadas de expansão e inflação atingi-se a convergência do algoritmo. As tabelas 9 a 10 a seguir ilustram algumas dessas etapas.

Tabela 9: Algumas Iterações do algoritmo MCL (após as operações $\Gamma_2(\Gamma_2 M^2 \cdot \Gamma_2 M^2)$).

0,448	0,080	0,023		0,068	0,426	0,359			0,432		
0,018	0,285	0,228	0,007	0,176	0,006	0,033	0,005		0,007		
0,005	0,223	0,290	0,022	0,173		0,010	0,017	0,003	0,001	0,003	0,001
		0,018	0,059	0,222	0,040		0,001	0,187	0,139		0,139
0,027	0,312	0,314	0,028	0,439	0,005	0,054	0,022	0,003	0,010	0,003	0,001
0,116	0,007	0,001		0,004	0,157	0,085				0,131	
0,096	0,040	0,013		0,037	0,083	0,197	0,001			0,104	
		0,012	0,042	0,172	0,029		0,002	0,198	0,133		0,133
		0,001	0,015	0,256	0,009			0,266	0,326		0,326
0,290	0,021	0,002		0,017	0,323	0,260				0,316	
		0,001	0,015	0,256	0,009			0,266	0,326		0,326
			0,001	0,037	0,001			0,039	0,069		0,069

Fonte: MACROPOL, 2009, p.30.

Até a convergência, onde a seguinte matriz é obtida:

Tabela 10: Matriz-resultado do algoritmo MCL.

1,000			1,000	1,000		1,000	
	1,000	1,000		1,000			
			0,500		0,500	0,500	0,500
			0,500		0,500	0,500	0,500

Fonte: MACROPOL, 2009, pp. 31-32.

Assim, pode-se ler diretamente a matriz-resultado para se identificar os *clusters* encontrados pelo algoritmo. Deste modo, neste exemplo identificam-se os seguintes *clusters*:

- *Cluster 1*: composto pelos vértices 1, 6, 7, 10;
- *Cluster 2*: composto pelos vértices 2, 3 e 5 e,
- *Cluster 3*: contendo os vértices 4, 8,9, 11 e 12.

Vale observar que a penúltima linha da matriz acima representa o mesmo *cluster 3*, podendo ser desconsiderada.

2.2.4 Outros algoritmos de clusterização em grafos

2.2.4.1 *K-means* de vetores de alta dimensão

Algoritmos do tipo *K-means* podem utilizar também os vetores construídos com o algoritmo desenvolvido nesta tese para codificação dos vértices de um grafo em tais estruturas. Deste modo, podem-se usar também estes algoritmos de clusterização em problemas de clusterização em grafos.

2.2.4.2 Heurísticas e Meta-heurísticas na clusterização em grafos

Soluções heurísticas de clusterização são utilizadas quando métodos tradicionais (não heurísticos) são muito lentos e de elevada complexidade algorítmica e se deseja então se encontrar uma boa solução aproximada. O objetivo de uma heurística é produzir uma boa solução para o problema.

Quando um problema de clusterização é modelado como um problema de otimização, as heurísticas encontram boas soluções viáveis para os mesmos. Como visto anteriormente, muitos problemas de clusterização em grafos são NP-ÁRDUO e abordagens com heurísticas e meta-heurísticas na abordagem de tais problemas passam a ser muito atrativas.

Diversas heurísticas surgiram visando a determinação de *clusters* em grafos.

Exemplos de heurísticas desenvolvidas para determinação de clusters em grafos são os trabalhos de GERASOULIS e YANG (1991), KIM e BROWNE (1988) e SARKAR (1989).

Abordagens com meta-heurísticas são formas de se encontrar boas soluções viáveis a problemas de otimização (normalmente na área de otimização combinatória), geralmente em problemas para os quais não se conhece algoritmos eficientes. Estas soluções utilizam combinações de escolhas aleatórias e conhecimento histórico dos resultados anteriores adquiridos (ao longo da execução desses algoritmos) para se guiarem e realizarem buscas locais em vizinhanças dentro do domínio da função sendo otimizada.

Deste modo, obviamente, abordagens deste tipo em problemas de clusterização demandam que o mesmo seja expresso como um problema de otimização.

Estes tipos de soluções para problemas de clusterização em grafos também tem sido vastamente empregados na literatura. Exemplos de abordagens destes problemas com meta-heurísticas são apresentados abaixo:

- SELIM, et all (2014) apresentaram uma abordagem baseada em Algoritmos Genéticos para a definição de *clusters* em grafos em um problema clusterização de grafos. Neste trabalho, os autores trataram do problema específico de clusterização em grafos (desenvolvido no âmbito da solução de um problema de *design* ótimo de *softwares* visando sua migração do paradigma estruturado para o paradigma orientado a objeto) como um problema de otimização quadrática inteira, a ser resolvido por este tipo de meta-heurística.

Deste modo, definiram como função-objetivo do problema de otimização (associado ao problema de clusterização) a maximização de arestas intra-*clusters*.

HINNE e MARCHIORI (2011) desenvolveram um algoritmo baseado na meta-heurística colônia de formigas para detectar cortes balanceados em grafos. Com estes cortes *clusters* são identificados.

2.2.4.3 Spectral Clustering

Em clusterização espectral de grafos, se utiliza os autovalores da matriz de similaridade dos dados de um grafo para se implementar uma redução de dimensionalidade antes da clusterização se dar, em dimensões menores. Uma referência introdutória ao assunto se encontra em CHARLES (2012).

2.2.4.4 Clusterização utilizando elementos de Teoria de Grafos

Alguns trabalhos de clusterização em grafos tem se utilizado de alguns elementos de Teoria de Grafos para a determinação de clusters. Os mais comuns são o uso de algoritmos de fluxo máximo/corte mínimo e o conceito de cliques em um grafo.

2.2.4.5 Clusterização baseada em cliques em grafos

Em Teoria de Grafos uma Clique (em um grafo não orientado) é um subconjunto de vértices que, cada dois vértices pertencentes ao mesmo conjunto são conectados por uma aresta. Um exemplo de clusterização utilizando o conceito de Clique é o trabalho de CHUI, ZHOU, QU, WONG e LI (2008).

Este conceito pode ser utilizado na identificação de *clusters* em um grafo.

2.2.4.6 Algoritmo de Fluxo Máximo/Corte Mínimo.

Algoritmos de fluxo máximo (e corte mínimo) podem ser executados repetidamente (para cada par de vértices do grafo) visando-se auxiliar na determinação de *clusters*. Para uma revisão de alguns algoritmos de fluxo máximo vide NETTO (2006).

2.3 ALGUMAS APLICAÇÕES ANTERIORES DE CLUSTERIZAÇÃO ENVOLVENDO TRAÇOS DE PERSONALIDADE

O objetivo desta sessão é fazer um breve apanhado de aplicações de clusterização feitas anteriormente em problemas envolvendo traços de personalidade.

Como já apresentado anteriormente, os primeiros trabalhos envolvendo clusterização de dados em Psicologia foram os trabalhos de ALLPORT E ODBERT (1936), ZUBIN (1938), em 1939 por TRYON (1970) e CATELL (1943a e b).

ALLPORT E ODBERT (1936) classificaram os termos descritores de personalidade presentes no dicionário Webster da época, em quatro categorias amplas que os autores batizaram de: *Neutral Terms Designating Possible Personal Traits* (termos neutros designando possíveis traços pessoais), *Terms Primarily Descriptive of Temporary Moods or Activities* (termos primariamente descritores de humores temporários ou atividades), *Weighted Terms Conveying Social or Characterial Judgments of Personal Conduct, or Designating Influence on Others* (termos concebendo julgamentos de cunho social ou de caráter com respeito à conduta pessoal ou ainda designando influência sobre outros) e *Miscellaneous: Designations of Physique, Capacities, and Developmental Conditions; Metaphorical and Doubtful Terms* (categoria envolvendo designações de condições física, de capacidade e de desenvolvimento; assim como termos metafóricos e que expressam dúvidas). Como se verá mais adiante, essa classificação inicial será depois utilizado em outros estudos psicológicos, inclusive no estudo de CATELL (1943a e b) que também se utilizou de técnica de clusterização de dados.

Esta classificação levou em consideração a experiência dos autores e foi levada a cabo por uma equipe contendo pessoas da área de Literatura e de Psicologia. Como todo

problema de classificação é também um problema de clusterização esse foi o primeiro grande estudo de clusterização envolvendo traços de personalidade (contudo o procedimento de clusterização se deu envolvendo conhecimentos dos autores na área e não um algoritmo quantitativo).

ZUBIN (1938) utilizou pela primeira vez na Psicologia um esquema quantitativo para determinação de *clusters* de indivíduos, que mais se aproximaria de um algoritmo fechado. Neste esquema de classificação de indivíduos em grupos (quanto a um determinado traço de personalidade pré-especificado) se utilizou principalmente do grau de concordância de *scores* obtidos em respostas a quesitos em questionários. Este trabalho de análise de *clusters* foi pioneiro como método de clusterização na Psicologia.

Em 1939, TRYON (1970) apresentou pela primeira vez na área de traços de personalidade um algoritmo formal de clusterização baseado em análise fatorial.

CATELL (1943b) também se utiliza de métodos de clusterização para reduzir um conjunto inicial de traços. Estes *clusters* foram determinados com base nas correlações entre os *ratings* obtidos em respostas a itens de questionários.

2.3.1 Outros estudos de clusterização envolvendo traços de personalidade a citar

Após estes estudos pioneiros utilizando métodos de clusterização, uma série de trabalhos e desenvolvimentos em Psicologia (e também associados à personalidade) tem surgido na literatura. A seguir se apresentará sucintamente alguns desses estudos.

- ARSHAVA, NOSENKO, NOSENKO E SAYEVICH (2013): esses autores desenvolveram um estudo onde abordam as diferenças individuais na auto-evolução de suas personalidades como refletindo a interfluência de situações específicas da vida real e os (novos) traços (contextualizados) de personalidade. Mais particularmente estes autores detectaram estas diferenças intra-*clusters* de indivíduos utilizando o algoritmo *K-means*.

3 CONSIDERAÇÕES SOBRE MÉTODOS DE CLUSTERIZAÇÃO DE PALAVRAS EM *THESAURI*

Neste capítulo se apresentará as principais considerações que devem ser feitas na abordagem de um problema de clusterização de palavras em um *thesaurus*. Este capítulo tem, como um dos seus principais objetivos destacar algoritmos de clusterização como métodos sistemáticos de organização dessas palavras segundo os esquemas propostos (*clusters*).

O primeiro arquiteto de um *thesaurus* amplamente conhecido foi Peter Mark Roget (1852), que definia como a principal finalidade de tal dispositivo ajudar o usuário a encontrar palavras pelas quais uma ideia pode ser mais bem expressa (ROGET, 2013).

Um *thesaurus* pode ser entendido como toda listagem que apresenta palavras relacionadas entre si de acordo com a similaridade de significados que partilham. Diferentemente de um dicionário, um *thesaurus* não define termos, mas os apresenta segundo esta organização. Assim, de acordo com BAKER e MCCALLUM (1998) um *thesaurus* organiza palavras semanticamente relacionadas.

Deve-se inicialmente ressaltar sucintamente aqui que o problema de determinação de *clusters* de palavras em um repositório de palavras deste tipo tem inúmeras aplicações práticas na literatura, indo desde o desenvolvimento de métodos de organização do conhecimento por similaridade do conteúdo de textos (como por exemplo, na esquemática proposta por PARVIN, MINAEI-BIDGOLI e DAHBASHI, 2012) desenvolvimentos de mecanismos de buscas na *internet* (a esse respeito vide BRASLAVSKI, ALSHANSKI e SHISHKIN, 2004) até aplicações em TP, como será aqui apresentado.

Estes esquemas de organização têm como principais características:- Ser sistemático significa que todas as etapas de organização das mesmas em *clusters* são conhecidas e reproduzíveis em qualquer conjunto de palavras.

- É um método alternativo de organização dessas palavras que evita possíveis coletas exaustivas de palavras de dicionários físicos ou outras bases de dados.

- Fornece uma organização ao menos inicial de palavras agrupadas segundo ideias/conceitos gerais.

Estas características fornecem aos potenciais usuários desses algoritmos um método atrativo de organização de palavras contidas em *thesauri*; especialmente quando se lida com grandes quantidades de palavras e se deseja uma organização das mesmas para uma posterior análise e eventual reorganização das mesmas, segundo outros critérios.

Neste capítulo se fará uma breve cobertura de algoritmos que podem ser empregados na determinação de *clusters* em grafos de palavras presentes no *thesaurus* da TT Software, na versão em planilha Excel, utilizado neste trabalho.

Deve-se iniciar esta sessão com a constatação de que todo *thesaurus* fornece ao seu usuário uma listagem de palavras relacionadas entre si.

Nestes *thesauri*, apresenta-se uma listagem de palavras relacionadas entre si, segundo determinados sentidos de associação. Em grande parte destas associações entre palavras, os sentidos de associação unindo pares de palavras relacionadas conferem às mesmas o papel de sinônimas umas das outras (no caso específico do *thesaurus* utilizado – segundo os próprios linguistas que o prepararam – a grande maioria das palavras vizinhas umas das outras cumprem o papel de sinônimas), como pode ser constatado a partir da carta enviada por linguista responsável pelo preparo do TTT, na tabela 1 do anexo.

Esta tendência decorre da própria natureza de todo *thesauri*, segundo os próprios linguistas que preparam o *thesaurus* adquirido – vide a carta do linguista responsável pelo preparo da versão do TTT utilizado, na tabela 1 no Anexo deste trabalho para os desenvolvimentos deste trabalho, na grande maioria dos casos, o relacionamento direto entre palavras neste repositório representa uma relação de sinônimo, ou seja, de palavras muito proximamente relacionadas entre si, segundo determinados sentidos de associação. Assim, palavras em um mesmo cluster, mesmo que indiretamente relacionadas, tendem a serem mais próximas do que em palavras pertencentes a clusters diferentes.

Deve-se ressaltar aqui que sinônimos são definidos como palavras com o mesmo significado ou significados similares.

Assim, os sinônimos são identificados em função de certos sentidos que as palavras podem receber. Assim, uma palavra que foi originalmente criada para designar algo, pode, com o passar do tempo, por exemplo, ganhar outros significados e desta forma passar a ser sinônimas de outras palavras, dentro de determinados sentidos.

Mais ainda, *clusters* de palavras neste repositório podem ser definidos de forma a propiciar que, palavras em um mesmo *cluster* apresentem um sentido ou ideia comum a todas elas isto é, um conceito – ou conceitos – ou ideia mais geral que se aplica a todas as palavras neste *cluster*. Deste modo, tem-se um meio de se organizar as palavras nestas bases de palavras de forma a garantir que estas palavras assim organizadas tenham sentidos ou conceitos mais gerais, que se apliquem a elas, desde que pertençam a um mesmo *cluster*.

Por exemplo, HIGGINSON e PHLEAN (1984), reconhecem e definem um *cluster* de palavras em um *thesaurus* como um grupo de palavras que compartilham entre si um ou mais conceito(s) global (is), mais abrangente(s), ilustrativamente, citam, por exemplo, que palavras –em inglês – pertencentes a um cluster formado por *house, cabin, mansion, e shack* são similares no sentido de que elas todas representam formas de abrigo.

Obviamente, esta(s) ideia (as) geral (is) representativa(s) de cada *cluster* não necessariamente reproduzem com exatidão todos os significados e interpretações possíveis que cada palavra, em cada *cluster*, pode assumir. Entretanto esta(s) ideia (as) geral (is) fornecem uma descrição geral, uma *rich Picture*, dos conceitos contidos em cada um destes *clusters*.

Esta propriedade pode ser facilmente constatada pela visualização de redes de relacionamento de palavras em *thesaurus*, como por exemplo, as disponibilizadas nos *sites* de *visual thesaurus*, disponíveis em <http://www.thesaurus.com> e <http://www.visualthesaurus.com/>, que correspondem às representações gráficas, *online*, de cadeias de relacionamentos das palavras mais proximamente relacionadas a uma determinada de palavra de interesse. Estas mesmas propriedades também podem ser verificadas nas motivações que levaram ao desenvolvimento de mecanismos de buscas na *internet* baseados em *clusters* de palavras em repositórios deste tipo (como por exemplo, no trabalho de BRASLAVSKI, ALSHANSKI e SHISHKIN, 2004).

Deste modo, embora se tenha sentidos de associação definidos para cada par de palavras vizinhas em cada *cluster*, existem sentidos mais amplos que unem palavras relacionadas indiretamente nestas partições do grafo que representa as associações entre as mesmas.

As palavras utilizadas para expressarem as ideias gerais de cada *cluster* podem ser escolhidas a partir do *thesaurus* inteiro, (HIGGINSON e PHLEAN, 1984). Neste último caso, serão escolhidas dentre as palavras pertencentes ao próprio cluster considerado. Seguindo este *framework*, poder-se-ia, por exemplo, estabelecer como uma palavra candidata inicial para melhor representar a ideia geral de um dado *cluster* aquela que possui o maior número ligações com os termos pertencentes ao mesmo.

No caso de um *cluster* conter mais de uma ideia geral, poder-se-ia escolher como tentativas iniciais de termos representativos dessas ideias, os vocábulos com maior número de ligações como os demais dentro do mesmo *cluster*.

Este trabalho utilizará conceito semelhante de *cluster* em *thesaurus*, admitindo ainda que, em cada *cluster* se possa identificar mais de um sentidos ou conceitos globais relacionando as palavras neles contidos. Vale ressaltar que, uma vez adotado este conceito de *cluster* e que se tenha organizado todas as palavras de um *thesaurus* (ou partes dele) em *clusters* assim definidos é imediato reconhecer que as fronteiras entre estes *clusters* não são exatas no sentido de que exista uma única escolha possível para a determinação destas fronteiras entre *clusters*.

Assim, os limites de cada *cluster* podem variar de autor para autor, ou mesmo de aplicação para aplicação e de *thesaurus* para *thesaurus*. No caso desses *clusters* serem determinados por algoritmos, é a lógica intrínseca de cada um deles que irá determinar estes agrupamentos e por conseqüente as palavras que pertencerão a cada um deles.

De qualquer forma, visando respeitar esta definição de *clusters* em *thesauri* se espera que os agrupamentos encontrados por algoritmos sejam suficientemente grandes para conterem alguma variabilidade de sentido e de palavras em cada um deles, mas ao mesmo tempo não grande demais para que as palavras neles contidas possam ser associadas a alguns conceitos mais amplos, que caracterizam cada *cluster*.

Ainda com relação à definição de *clusters* em grafos, SCHAEFFER (2007) pondera o seguinte, em relação às propriedades desejáveis de tais agrupamentos:

Unfortunately, no single definition of a cluster in graphs is universally accepted, and the variants used in the literature are numerous. In the setting of graphs, each cluster should intuitively be connected: there should be at least one, preferably several paths connecting each pair of vertices within a cluster.

Assim, entendendo-se o conjunto de relacionamentos entre palavras de um *thesaurus* como um grafo, os *clusters* dessas palavras devem respeitar as ponderações feitas acima.

Assim sendo, *clusters* de palavras, nestes repositórios, contendo somente traços de personalidade, também partilham das mesmas características. Neste sentido, para obterem-se tais *clusters*, por emprego de algoritmos de clusterização, devem-se aplicar tais métodos a subconjuntos desta base dos dados que contenham somente estes tipos de palavras ou traços. Neste esquema, os traços estariam organizados segundo o que a literatura chama de *similarity trait clusters* – STC – ou *clusters* de similaridade de traços.

Quando se concebe a utilização de métodos de clusterização de palavras em *thesaurus* com vistas a se elaborar um método de organização destas palavras, seria desejável também que os *clusters* encontrados ou o maior número possível deles gozassem da característica de possuírem alto grau de conectividade intra *clusters* e baixo grau de conectividade inter *clusters*.

Agrupamentos de palavras com estas características tornam-se especialmente atrativos para o propósito de fornecerem uma organização destes termos; haja vista que em cada um deles se estaria agrupando termos estritamente relacionados entre si (de acordo com a linguagem natural) ao mesmo tempo em que palavras em *clusters* diferentes tendem a ser menos relacionadas.

Haja vista que se pode caracterizar e identificar uma determinada cultura de um determinado povo por meio das palavras que utilizadas para representar categorias do mundo real; assim como os diferentes sentidos que estas palavras podem ganhar em diferentes situações, se relacionando com outras o *thesaurus* é uma excelente forma de representação desta cultura.

Ademais, como vocábulos surgem e perdem uso ao longo do tempo, assim como os sentidos que assumem o *thesaurus* pode ser utilizado como uma expressão dinâmica da forma como as sociedades humanas escolhem vocábulos para atribuir sentidos (e, portanto escolhem também os sentidos comportamentais e ligados à personalidade contida nesses mesmos vocábulos). Neste contexto, seria extremamente desejável se desenvolver uma ferramenta que cumpra as funções anteriormente mencionadas em qualquer *theaurus* que fosse atualizado ao longo do tempo.

Outro aspecto relevante envolvendo aplicações com palavras em *thesaurus* refere-se à decisão dos tipos de palavras a serem incluídas no estudo. Como visto anteriormente o *thesaurus* utilizado neste trabalho lista, para cada relacionamento entre cada par de palavras, a função gramatical que um determinado termo cumpre ao se associar com uma determinada *keyword*, variando entre categorias como nome, verbo, adjetivo, dentre outras.

Deste modo, pode-se escolher que palavras, cumprindo que tipos de funções gramaticais, deveriam ser incluídos em uma determinada aplicação.

No caso específico da determinação de agrupamentos de traços de personalidade, os estudos psicoléxicos anteriores se detiveram às palavras cumprindo classes de nomes e adjetivos e por esta razão, profissionais da área e pesquisadores provavelmente irão usar na grande maioria dos casos palavras de bancos de dados que cumpram essas classes gramaticais.

Alguns autores incluem também, expressões inteiras para designar um traço de personalidade, como exemplo tem-se o trabalho de CATELL (1943b) em que o mesmo utiliza expressões como *special abilities* para caracterizar alguns traços. Mesmo nestes casos, com o auxílio de um psicólogo, seria possível se alterar o *thesaurus* original e incluir estas

expressões neste repositório de palavras, se determinando com quais outras palavras estas expressões devem estar associadas.

Como se verá ao longo deste trabalho, pode-se representar as palavras em um *thesaurus* e as relações entre as mesmas por meio de um grafo direcionado ou não. Neste esquema, cada palavra é representada por um vértice deste grafo, e seu relacionamento com outras palavras por arestas ou arcos, no caso de um grafo direcionado.

No caso da representação por um grafo direcionado, a relação entre duas palavras deve ser sempre representada por um par de arcos em sentidos opostos; isto para garantir que, independentemente de onde se parta, no grafo, para cada par de palavras diretamente relacionadas entre si, se tenha caminho entre elas, no grafo representando a rede de relacionamentos entre as palavras selecionadas do *thesaurus*.

Outro aspecto que se deve ter em mente, quando se aplica algoritmos de clusterização em grafos representando relacionamentos de palavras em *thesaurus* é que, os *clusters* encontrados refletem a própria estrutura de ligações ou relacionamentos entre as palavras consideradas segundo a própria estrutura de ligações presente neste banco de palavras isto é, segundo as próprias relações semânticas entre as mesmas, identificadas pelos linguistas que elaboraram este banco. Neste sentido, deve ressaltar aqui que se está ciente de que, em aplicações deste tipo pode ocorrer alguma discordância sobre se determinada(s) palavra(s) deveria(m) estar (em) diretamente ligada(s) a outra(s) palavra(s) ao invés da(s) palavra(s) a que efetivamente estão diretamente associadas no *thesaurus* adquirido. A seguir se explicará melhor este ponto.

3.1 NOTAS IMPORTANTES SOBRE A ESTRUTURA DE RELAÇÕES ENTRE PALAVRAS EM UM *THESAURUS*

Cabe aqui ressaltar de que se deve estar ciente de que algumas associações entre palavras no *thesaurus* podem, em alguns casos e para alguns leitores, não serem de imediato reconhecimento ou de fácil aceitação por parte dos mesmos; mas nestes casos isto se deve ao fato de que são essas as ligações presentes neste banco de palavras e, portanto o entendimento dos linguistas que prepararam o mesmo; assim sendo, existindo um sentido que liga tais palavras mesmo que este sentido não seja de fácil identificação.

Isto é, deve-se reconhecer a possibilidade de que algumas associações entre palavras presentes no *thesaurus* podem não ser imediatas ou mesmo aceitas por especialistas da Psicologia e demais áreas que possam se interessar em *clusters* de palavras. Contudo tais dificuldades podem ser contornadas pela consulta a estes linguistas ou, em última instância e

o *thesaurus* utilizado permite isso a alteração manual das associações entre estas palavras, neste banco de palavras no caso do *thesaurus* da TT Software, isto pode ser facilmente feito por meio da edição das células em planilha Excel que contém as palavras em questão antes que se possam aplicar métodos sistemáticos na busca de clusterização de palavras relacionadas.

Como corolário dessa possibilidade levantada, pode ocorrer, por parte do pesquisador que programou os algoritmos de clusterização aqui empregados, que alguns dos clusters por eles encontrados, contenham palavras cujos relacionamentos não sejam de imediato reconhecimento. Nestes casos, o *thesaurus* deverá ser alterado na versão utilizada neste trabalho este procedimento é fácil, bastando, para tanto, se alterar as posições das palavras na planilha Excel de forma a atender os entendimentos do usuário para que reflita um relacionamento de palavras relevante ao estudo e novamente se proceder à aplicação dos algoritmos de clusterização.

Outro aspecto que deve ressaltado é que, devido ao próprio desenvolvimento constante de qualquer língua, pode-se deparar com a necessidade de se incluir mais relacionamentos entre pares de palavras para se dar conta das diversas conotações que as mesmas vão ganhando. Novamente isto não é um obstáculo para a aplicação dos métodos de clusterização apresentados a seguir, bastando novamente se proceder a alterações do *thesaurus* original, segundo entendimento do especialista ou pesquisador.

Talvez o aspecto mais importante a ser esclarecido no que se refere ao TTT adotado é o significado das relações entre os pares de palavras, no mesmo *cluster*. Segundo os próprios linguistas que prepararam esta base de palavras utilizada, a grande maioria das relações entre as palavras presentes no *thesaurus* em questão representam relações semânticas entre palavras que cumprem papel de sinônimos, vide carta do linguista da empresa *True Term Thesaurus Database*, responsável pelo reparo do *thesaurus* utilizado, na tabela 41 do Anexo.

Por fim, esses linguistas também detectaram um caso onde a associação entre um determinado grupo de palavras não cumpre exatamente nem um papel de antônimo, nem de sinônimo, mas sim um papel explicativo, enumerando tipologias de um determinado conceito. Este caso é apresentado a seguir.

3.1.1 Exemplos de que nem todas as associações cumprem funções de sinônimos ou antônimos

A seguir se apresenta dois blocos de associações entre palavras no banco de palavras utilizado:

activity [v]: *grow; develop; age; ripen; evolve; mature; mellow*

property [a]: *adult; developed; grown*

No primeiro bloco, verifica-se que *grow, develop, age, ripen, evolve, mature e mellow* são formas de atividade (*activity*). Assim, estas palavras não são propriamente sinônimas de *activity*, mas tipologias desta última.

De modo similar, no segundo *bloco*, *adult, developed e grown* são tipologias de *property* (propriedade).

Desta forma, a esmagadora maioria das associações entre palavras, no *thesaurus*, representam relações semânticas onde as palavras estão associadas, dentro de determinados sentidos que lhes conferem o status de sinônimas, salvo raras exceções. Salvo raras exceções, as palavras incluídas em um mesmo *cluster*, segundo qualquer uma das metodologias apresentadas a seguir, são palavras mais fortemente relacionadas segundo determinados sentidos de associação entre elas.

4 ALGUNS DOS PRINCIPAIS DESENVOLVIMENTOS NA ÁREA DE TRAÇOS DE PERSONALIDADE ANTERIORMENTE FEITOS

4.1 CONCEITOS E CONCEPÇÕES ACERCA DE TRAÇOS DE PERSONALIDADE

A ideia de traços de personalidade, segundo MATTHEWS (2003), pode ser tão velha quanto à linguagem humana em si. De acordo com este autor:

The idea of personality traits may be as old as human language itself. Aristotle (384–322 BC), writing the Ethics in the fourth century BC, saw dispositions such as vanity, modesty and cowardice as key determinants of moral and immoral behaviour. He also described individual differences in these dispositions, often referring to excess, defect and intermediate levels of each. His student Theophrastus (371–287 BC) wrote a book describing thirty ‘characters’ or personality types, of which a translator remarked that Theophrastus’s title might better be rendered ‘traits’ (Rusten, 1993). Basic to his whole enterprise was the notion that individual good or bad traits of character may be isolated and studied separately. (p.3)

As concepções do dia-a-dia acerca de traços de personalidade (TP) se baseiam em duas hipóteses centrais:

Primeiro, que os TP são estáveis ao longo do tempo, embora assumam que o comportamento de um indivíduo possa variar em algum grau de ocasião para ocasião, assumem que existe uma consistência central neste comportamento. Assumem assim, que existem diferenças entre indivíduos que são aparentes e reconhecíveis ao longo de diferentes situações.

Geralmente assumem que os traços influenciam diretamente o comportamento.

Estas concepções, de acordo com MATHEWS (2003) enfrentam em boa parte o problema da circularidade existente entre traços e comportamentos. Como ressalta este autor, ARISTÓTELES (1104ac) sugeriu uma hipótese causal mais sutil, recíproca: que é através das ações que as disposições se desenvolvem; disposições estas que por sua vez influenciam ações. Mais especificamente sugere que: “It is by refraining from pleasures that we become temperate, and it is when we have become temperate that we are most able to abstain from pleasures (THOMSON, 1976, pp. 33–35).

Deste modo destaca que uma das principais tarefas da Psicologia científica é distinguir as propriedades internas das pessoas dos comportamentos observados e identificar o relacionamento entre os dois. Para se evitar esta circularidade é essencial que se identifique

as bases fisiológica, psicológica e social dos traços, que proporcionam as verdadeiras causalidades no comportamento observado.

4.1.1 Concepções científicas de TP

Conforme salienta MATHEWS (2003) existem diferenças significativas entre a concepção e traços apresentada acima e uma concepção científica dos mesmos. De acordo com o autor, diversas etapas devem ser cumpridas para se desenvolver uma ciência de TP:

A primeira etapa é a referente à medida e classificação dos TP. Neste sentido, a técnica mais simples (e utilizada) é se perguntar ao indivíduo sendo analisado ou alguém próximo a ele o quão bem determinados adjetivos-traço se aplicam a ele, por meio de aplicações de *ratings* a cada adjetivo. Alternativamente pode-se questionar estas pessoas sobre determinados comportamentos que se relacionam à personalidade, por exemplo, para se tentar inferir o grau de extraversion de um indivíduo pode-se perguntar a alguém que o conheça bem sobre seu comportamento social, a quantos clubes e sociedades pertence etc.

Alternativamente aos relatórios verbais e *ratings* em questionários pode-se tentar medir traços por meio de comportamentos observados em laboratório.

De acordo com MATHEWS (2003), os relatórios verbais tem sido o método principal de abordagem e medição dos traços.

Como existem muitas palavras descritoras da personalidade, muitas delas possuem significados similares; definindo TP que se sobrepõem uns aos outros quanto a esses significados. Tais TP podem ser agrupados de forma a definirem aspectos, ou dimensões, amplos da personalidade. Deve-se observar aqui que é também neste contexto que algoritmos de clusterização de dados, aplicados a TP relacionados segundo a linguagem natural, fornecem uma organização inicial dos mesmos, o que pode auxiliar o pesquisador na determinação de tais dimensões.

A segunda etapa se refere ao teste de como determinados traços se relacionam com comportamentos, por exemplo, se verificando como os *ratings* obtidos em questionários se correlacionam com medidas objetivas de comportamento. A hipótese implícita na abordagem por traços de personalidade é que as pessoas tendem a se comportar consistentemente em diferentes situações.

A última etapa se refere a um desenvolvimento de uma teoria capaz de, para cada traço pertencente ao modelo e a partir de determinados níveis especificados para cada um deles prever alguns aspectos relevantes do comportamento de um indivíduo.

Neste sentido não basta medir os diferentes níveis de avaliação que os indivíduos são avaliados segundo um conjunto de traços e verificar seu comportamento, há a necessidade da elaboração de uma teoria que concilie os diferentes níveis de avaliação segundo esses traços, o comportamento observado com explicações consistentes de como essas associações se dão e de porque esses conjuntos de traços preveem comportamento.

Uma dificuldade para a construção dessas teorias é que a personalidade pode ser representada por uma variedade de níveis de descrição psicológica. Como exemplifica MATTHEWS (2003), p.5, extraversão pode estar associada a propriedades do sistema nervoso central, ou com crenças e conhecimento social adquirido.

Segundo este autor, só se pode distinguir essas possibilidades por métodos científicos de formulação de hipóteses específicas e com o teste das mesmas frente a evidências observacionais.

O desenvolvimento de modelos e teorias acerca da personalidade tem se baseado em algumas hipóteses geralmente aceitas. Na concepção da teoria de traços tradicional se verifica o respaldo em duas hipóteses centrais para a elaboração de explicações acerca do comportamento humano: a primazia causal e a hipótese do *inner locus*, a seguir descritas.

- **Primazia Causal**. Embora exista incerteza sobre o status causal dos traços e que possa existir reciprocidade entre a influência causal dos traços e comportamentos observados, se assume frequentemente a hipótese de primazia causal, que alega que a direção dominante de causalidade é dos traços para o comportamento observado.

Assim, supondo-se que um indivíduo tenha recebido um *score* elevado de neuroticismo e que também apresente sintomas clínicos de depressão. Com relação a este quadro, Poder-se-ia conceber uma das seguintes possibilidades:

O neuroticismo causou a depressão.

A depressão causou o neuroticismo.

Depressão e neuroticismo foram independentemente influenciadas por um terceiro fator.

Adotando-se esta hipótese se estaria assumindo que a opção a) é a correta ou ao menos que a direção dominante de causalidade é a dos traços para os comportamentos observados.

Deste modo, embora medidas de traços tais como as obtidas por *scores* de questionários, não sejam agentes causais em si, elas indexam estruturas fisiológicas e psicológicas (a serem estabelecidas por uma teoria da personalidade) que diretamente influenciam comportamento.

Deve-se destacar aqui a pluralidade de concepções dos atributos que uma teoria causal de traços de personalidade na literatura. ALLPORT e ODBERT (1936) concebiam traços como estruturas mentais organizadas que iniciavam e guiavam o comportamento.

Na concepção de HETTEMA e DEARY (1993), a explicação do comportamento requer diferentes tipos de análises, incluindo genética, fisiológica e da influência de fatores de aprendizagem e sociais.

Deste modo modelos descritores das causalidades dos traços no comportamento variarão dependendo do nível investigado (embora no entender de MATHEWS, 2003, uma teoria final de traços de personalidade deveria inter-relacionar estes diversos níveis).

Também vale a pena se observar aqui que os efeitos (causais) dos traços no comportamento podem ser indiretos. Como observa MATHEWS (2003, p.6) traços interagem com fatores situacionais produzindo condições internas transientes, os quais por sua vez podem influenciar mais diretamente no comportamento que estes mesmos traços.

- *Inner Locus*: Essa hipótese é caracterizada pela suposição de que alguns traços de personalidade (ou mesmo algumas dimensões da personalidade) podem se relacionar com algumas qualidades mais fundamentais, mais centrais do indivíduo; e estas sim é que seriam mais importantes na descrição de algumas peculiaridades da sua personalidade.

Essas qualidades mais fundamentais, seriam potencialmente explicadas por fatores de diferentes naturezas: fisiológicos, genéticos ou cognitivos; constituindo assim as verdadeiras fontes ou *locus* mais central da causalidade de um determinado comportamento (*inner locus* desse comportamento).

Por exemplo, de acordo com MCCRAE (2000) traços importantes como extroversão e neuroticismo (que pode ser entendido como uma tendência ampla de se experimentar emoções negativas) podem ser bastante influenciadas por fatores genéticos.

Ainda como exemplo de como a noção de *inner locus* está presente em diversos de traços de personalidade, pode-se citar o trabalho de CATELL (1943a, 1943b), que distinguiram os traços de superfície (*surface traits*) dos traços de fonte ou origem (*source traits*); os primeiros sendo simplesmente respostas de traços mais profundos, capazes de explicar de forma mais consistente o comportamento.

4.2 REVISÃO DOS DESENVOLVIMENTOS EM TRAÇOS DE PERSONALIDADE

De acordo com MATTEWS (2003, p.8), o estudo científico dos traços formaliza a tendência da linguagem natural de usar traços como descritores de indivíduos. Ele também

formaliza o conhecimento popular de que existem generalidades no que se refere à personalidade humana; de tal sorte que indivíduos com disposições similares podem ser agrupados em categorias para fins de descrição dos diversos comportamentos. Neste sentido, para este autor, a história dos traços sofre influências da psicologia popular e da linguagem natural.

Esta história, segundo ele, pode ser contada de diversas formas: da busca de conceitos contrapostos aos traços extravertido e neuroticismo ao modelo atualmente dominante dos Cinco Fatores da Personalidade (GOLDBERG, 1993); sofrendo influências do pensamento clássico, passando pelos primeiros trabalhos científicos de traços até os modelos dos Cinco Fatores.

4.2.1 Condições necessárias para o desenvolvimento da pesquisa científica de traços

Para que os primeiros desenvolvimentos científicos com traços de personalidade pudesse se dar, algumas condições tiveram que ser satisfeitas; a saber:

- a disponibilidade de bases de dados e da coleção sistemática desses dados;
- o desenvolvimento de técnicas estatísticas de análise de dados (em particular o desenvolvimento de técnicas de correlação de dados e os primeiros desenvolvimentos de técnicas de análise fatorial) e
- o desenvolvimento de teorias passíveis de serem testadas.

No que tange ao desenvolvimento da análise estatística, foi somente com o desenvolvimento da análise fatorial que foi possível se reduzir o enorme número de termos-traços em um número manejável de dimensões amplas da personalidade. Foi a utilização destas técnicas que marcou o surgimento da pesquisa moderna nesta área.

4.2.2 Os primeiros estudos empíricos

Como se verá, os primeiros estudos na área se utilizaram de dicionários como fonte de termos traços. Estes estudos se apoiaram na hipótese léxica, que nas palavras de ALLPORT e ODBERT (1936)

Estes autores identificaram quase 18.000 termos relevantes à personalidade humana. Estes autores hipotetizavam que:

Those individual differences that are most salient and socially relevant in people's lives will eventually become encoded into their language; the more important such a difference, the more likely is it to become expressed as a single word.

Esta afirmativa ficou conhecida como Hipótese Léxica.

Também foi importante para os primeiros desenvolvimentos na área o reconhecimento da natureza preditiva dos traços (no sentido de que se sabendo os traços de um indivíduo se poderia prever seu comportamento futuro).

4.2.2.1 Estudo de Allport e Odbert (1936)

Allport and Odbert, como já mencionado, utilizaram esta hipótese para identificar traços de personalidade em um dicionário Webster na língua inglesa. Deste dicionário extraíram cerca de 18.000 palavras descritores de personalidade.

Estes autores acessaram a segunda edição do dicionário *Webster's New International Dictionary* em busca de potenciais descritores de personalidade. Com este intuito coletaram 17.953 termos que se aplicavam ao comportamento humano. Deste conjunto, extraíram 4.500 adjetivos descritores de personalidade, que consideraram descrever traços observáveis e relativamente permanentes.

A seguir estas palavras foram classificadas em quatro grupos, criados arbitrariamente por estes autores de acordo com suas principais finalidades: o grupo de traços pessoais, o de estados temporários, o de avaliações sociais e o de termos metafóricos ou que expressam dúvidas.

De acordo com esses autores:

The list that follows contains all the words descriptive of personality or personal behavior (save those that are obsolete) included in Webster's New International Dictionary. The edition used (1925) comprises approximately 400,000 separate terms or derivatives. Our list contains in all 17,953 words, or 4 1/2% of the total English vocabulary. A very few common terms, chiefly slang, have been added, but no attempt has been made to follow systematically any other source than Webster's Dictionary. The criterion for inclusion consists in the capacity of any term to distinguish the behavior of one human being from that of another. Terms representing common (non-distinctive) behavior are excluded, e.g., walking and digesting, whereas more differentiating and stylistic terms applied to these same activities, such as mincing and dyspeptic, are included. In many cases the application of this criterion involved a considerable degree of arbitrariness. In deciding doubtful cases the dictionary definition was followed: if in any of its meanings a term might be differentially employed in characterizing personal behavior it was admitted. Adjectival and participial forms have been preferred throughout; nouns and adverbs appear only where no corresponding adjective or participle exists, or else in cases where their meaning is distinctive (e.g., both Quaker and Quakerish are included). Various adjectival variants of a word are given only when the shades of meaning are readily distinguishable. A person may be sour because of innate temperamental constitution; the term soured implies some environmental genesis; and sourish indicates a difference in degree and perhaps a difference in kind. All three terms are included. With many other word-forms the differentiation is non-semantic: vagaries of morphological development rather than the need for precise distinctions in meaning are responsible for such

clusters as atrabilarian, atrabilarious, atrabiliar, atrabiliarious, atrabiliary, and atrabilious." ALLPORT e ODBERT, 1936, PP. 25-25

A tabela 11 a seguir contém a primeira das 134 páginas de termos captados nesse estudo.

Tabela 11: Tabela ilustrativa com a primeira das 134 páginas de termos captados no estudo de Allport e Odbert (1936).

<i>COLUMN I</i>	<i>COLUMN II</i>	<i>COLUMN III</i>	<i>COLUMN IV</i>
<i>Neutral Terms Designating Possible Personal Traits</i>	<i>Terms Primarily Descriptive of Temporary Moods or Activities</i>	<i>Weighted Terms Conveying Social or Characterial Judgments of Personal Conduct, or Designating Influence on Others</i>	<i>Miscellaneous: Designations of Physique, Capacities, and Developmental Conditions; Metaphorical and Doubtful Terms</i>
abandoned	abashed	abnormal	able
abject	ablaze	absorbing	abortive
abrupt	absent	absurd	abrasive
absent-minded	absorbed	abundant	absinthine
absolutist	abstracted	acceptable	absolute
absorbent	abusive	accidental	abstract
abstemious	accusing	acclaimed	abstractive
abstentious	acrimonious	accomplished	abysmal
abstinent		accountable	acanaceous
abstruse		accountless	acaroid
academic		accumbrous	accidental
accelerative		accursed	accipital
accendible		acknowledged	acentric
accessible		addle-brained	acerb
accommodating			acerbic
accordiable			acenscent
accostable			acetose
accrescent			Achatean
accretive			Acherontic
accroaching			Achilleian
accurate			achromic
acerbate			aciculate
acharné, F.			acid
acquiescent			acidiferous
acquisitive			acidulous
acrobatical			acquainted
active			acquired
acute			acranaial
adamantean			acrid
			acronarcotic
			adagio
			adaptive
			addicted
			Addisonian

Fonte: DE RAAD, 2000, p.4.

4.2.2.2 Estudos de Raymond Catell

De acordo com MATTHEWS (2003, p.18) a discussão de traços primários deve começar com Raymond B. Cattell; cujo projeto, de acordo com esse autor é:

...the most ambitious ever undertaken in psychology". It seeks to explain individual differences in every area of life from psychometrically sound measures of ability, motivation, personality and mood. Massive quantities of data have been generated by this enterprise, along with several widely used questionnaires and tests.

Catell começou suas pesquisas na área de traços com trabalhos envolvendo termos descritivos de comportamentos coletados em dicionários, mudando logo a seguir o foco de suas pesquisas para questões relacionadas com itens de questionários.

Nestes esforços identificou um modelo contendo vinte e três fatores primários da personalidade. As dezesseis dimensões mais robustas dessas dimensões são medidas pelo Questionário dos Dezesseis Fatores da Personalidade (16 PF *Questionnaire*), que se tornou na pesquisa e outras utilizações aplicadas por muitas décadas.

4.2.2.2.1 A identificação de 171 traços de personalidade com base em análise fatorial

Raymond Catell, usando como base o estudo de Allport e Odbert (1936) usou uma lista de 4.504 palavras, encontradas na primeira coluna da lista de palavras deste estudo, batizadas por estes autores de “termos de trato” e mais umas poucas centenas de termos adicionais; a maior parte deles originados da segunda coluna da lista deste estudo (coluna esta que rotulava os termos a ela pertencente de “estados temporários”).

Esta lista original foi subsequentemente filtrada e reduzida para um número de termos menor que a metade do número inicial, com a eliminação de termos com prefixos e de termos considerados arcaicos e coloquiais por Catell.

A seguir, ajuda de um estudante de literatura e um psicólogo, categorizaram as palavras, o que resultou em cerca de 160 categorias, com uma média de 13,4 sinônimos por categoria.

Dessas 160 categorias o autor finalmente elegeu 171 traços (comuns) de personalidade para fazerem parte de um questionário. Com base nas correlações obtidas nas respostas aos itens deste questionário aplicou análise de *clusters* para reduzir ainda mais esse conjunto em *clusters* (CATTELL 1943b).

Com um novo processo da seleção, chegou em 171 palavras, apresentadas a seguir.

Tabela 12: Os 171 termos derivados da primeira condensação feita no estudo de Cattell (1943b)

PERSONALITY VARIABLES COMPRISING THE COMPLETE PERSONALITY SPHERE	
FIRST CONDENSATION FROM UNABRIDGED DICTIONARY AND PSYCHOLOGICAL LITERATURE FIELDS	
TRAIT OR VARIABLE.....	OPPOSITE OF TRAIT OR VARIABLE (if Any)
1. ABILITIES. <i>Intelligence.</i>	Capacity to perceive relations, insight, quickness to learn, adaptability in problems.
2. " " <i>Special Abilities.</i>	<i>Drawing.</i> Facility in graphical representation.
3. " " " "	<i>Mathematical.</i> Thurstone's <i>N</i> or number ability.
4. " " " "	<i>Manual Dexterity.</i> (See Cox's <i>M</i> factor.)
5. " " " "	<i>Mechanical Aptitude.</i> Facility in constructing and understanding machinery.
6. " " " "	<i>Musical Aptitude.</i> (See Seashore.)
7. " " " "	<i>Physical Strength and Endurance.</i>
8. " " " "	<i>Logical Ability, Reasoning.</i> Thurstone's <i>I</i> .
9. " " " "	<i>Spatial, Visual Ability.</i> Thurstone's <i>S</i> and Kelley's factors.
10. " " " "	<i>Verbal Attitude.</i> Thurstone and Spearman's <i>V</i> factor. Facility in right use of words.
11. ALERT.....	ABSENT-MINDED
12. ACQUISITIVE.....	
13. AFFECTED.....	NATURAL
14. AFFECTIONATE.....	FRIGID
15. AGORAPHIC.....	

16.	ALCOHOLIC	
17.	AMBITIOUS	UNAMBITIOUS
18.	AMOROUS	LUSTLESS
19.	ANALYTICAL	
20.	ANTEVERT	RETROVERT
21.	ARGUMENTATIVE	
22.	ARROGANT	HUMBLE
23.	ASCETIC	SENSUOUS
24.	ASSERTIVE	SUBMISSIVE
25.	AUSTERE	PROFLIGATE
26.	AUTOCRATIC	
27.	BOASTFUL	MODEST
28.	BROODING	UNREFINING
29.	CAUTIOUS	RECKLESS
30.	CHARMING	
31.	CHEERFUL	GLOOMY
32.	CLEAR THINKING	INCOHERENT
33.	CLEVER	
34.	CLAUSTROPHOBIC	
35.	CONCEIVED	SELF-DISSATISFIED
36.	CONSCIENTIOUS	CONSCIENCELESS
37.	CONSTRUCTIVE	
38.	CONTENTED	DISSATISFIED
39.	CONVENTIONAL	INDIVIDUALISTIC
40.	COOPERATIVE	OBSTRUCTIVE
41.	COURAGEOUS	COWARDLY
42.	CURIOUS	UNENQUIRING
43.	CYNICAL	IDEALISTIC
44.	DEBONNAIRE	
45.	DEFENSIVE	
46.	DELIBERATIVE	DECISIVE
47.	EASY-GOING	SHORT-TEMPERED
48.	ECCENTRIC	
49.	EFFEMINATE	MASCULINE
50.	Egotistical	ALTOCENTRIC
51.	ELOQUENT	INARTICULATE
52.	EMOTIONAL I	UNEMOTIONAL
	Emotionality in all varieties of emotion. Burt's general emotionality factor.	
53.	EMOTIONAL II (Burt's Sthenic-asthenic factor)	UNEMOTIONAL
	Emotionality in sociability, assertive.	
54.	ENERGETIC-SPIRITED	LANGUID
55.	ENERGETIC-INDUSTRIOUS	INACTIVE-INDOLENT
56.	ENTERPRISING	SHIPTLESS
57.	ENTHUSIASTIC	APATHETIC
58.	EVASIVE	FACING LIFE
59.	EXCITABLE	PHLEGMATIC
60.	EXTRA-PUNITIVE	PRaisEful
61.	EXHIBITIONISTIC	SELF-EFFACING
62.	FAIR-MINDED	PARTIAL
63.	PASTIDIOUS	COARSE
64.	FLATTERING	
65.	FORMAL	CASUAL
66.	FRANK	SECRETIVE
67.	FRIENDLY	HOSTILE
68.	GENEROUS	TIGHT-FISTED
69.	GENIAL	COLD-HEARTED
70.	GLUTTONOUS	QUEASY
71.	GRATEFUL	THANKLESS
72.	HABIT-BOUND	LABILE
73.	HARD	SOFT-HEARTED
74.	HEADSTRONG	GENTLE TEMPERED
75.	HEARTY	QUIET
76.	HIGH-STRUNG	RELAXED

77.	HONEST	DISHONEST
78.	HURRIED	LETHARGIC
79.	HYPOCHONDRIACAL	
80.	IMAGINATIVE	DULL
81.	IMITATIVE	
82.	IMPULSIVE TEMPERAMENTALLY	DELIBERATE
83.	INDEPENDENT	DEPENDENT
84.	INFLEXIBLE (EMOTIONALLY)	ADAPTABLE (to change)
85.	INHIBITED	INCONTINENT
86.	INTERESTS WIDE	INTERESTS NARROW
87.	INTERESTS SPECIAL, <i>Aesthetic</i> (GENERAL)	
88.	" " <i>Artistic</i> (PAINTING, ART, ARCHITECTURE)	
89.	" " <i>Economic</i>	
90.	" " <i>Home and Family</i>	
91.	" " <i>Music</i>	
92.	" " <i>Physical Activity</i>	
93.	" " <i>Political</i>	
94.	" " <i>Religious</i>	
95.	" " <i>Social</i>	
96.	" " <i>Theoretical</i>	
97.	" " <i>Technical</i>	
98.	INTUITIVE	LOGICAL
99.	INTROSPECTIVE	
100.	IRRITABLE	GOOD TEMPERED
101.	JEALOUS	
102.	KIND (by disposition)	RUTHLESS
103.	KIND (on principle)	
104.	LAUGHTERFUL	MIRTHLESS
105.	LEADING (not Domineering)	
106.	LOYAL	FICKLE
107.	MATURE (in emotional development)	INFANTILE
108.	MEMORY GOOD	FORGETFUL
109.	MISCHIEVOUS	
110.	MULISH	REASONABLE
111.	MYSTICAL	APOLLONIAN
112.	NEUROTIC	
113.	OPINIONATED	TOLERANT
114.	OPTIMISTIC	PESSIMISTIC
115.	ORIGINAL	BANAL
116.	PATIENT	IMPATIENT
117.	PAINSTAKING	SLIPSHOD
118.	PEDANTIC	DISORDERLY
119.	PERSEVERING	QUITTING
120.	PHANTASYING	
121.	PHYSICALLY ACTIVE	
122.	PIOUS	WORLDLY
123.	PLAINITIVE	
124.	PLANFUL	PLANLESS
125.	POISED	AWKWARD
126.	POLISHED	ROUGH
127.	PRACTICAL	UNREALISTIC
128.	PUGNACIOUS	PEACEABLE
129.	RELIABLE	UNDEPENDABLE
130.	RESERVED	INTRUSIVE
131.	RESILIENT	DEPRESSIBLE
132.	RESPONSIVE	ALOOPE
133.	REVERENT	REBELLIOUS
134.	SADISTIC	MASOCHISTIC
135.	SARCASTIC	
136.	SELF-CONFIDENT	SELF-DISTRUSTING
137.	SELF-CONTROLLED	
138.	SELF-DECEIVING	
139.	SELF-PITYING	

140.	SELF-RESPECTING	
141.	SELFISH	SELF-DENYING
142.	SENSITIVE	TOUGH
143.	SENTIMENTAL	HARD-HEADED
144.	SERIOUS	FRIVOLOUS
145.	SHREWD	NAÏVE
146.	SLANDEROUS	
147.	SLEEPS WELL	SLEEPS POORLY
148.	SLOW (temperamentally, in movement, in reaction)	QUICK
149.	SOCIABLE I	SHY
	Fond of meeting people, good mixer.	
150.	SOCIABLE II	EXCLUSIVE
	Gregarious, congregative, companionable.	
151.	SOPHISTICATED	SIMPLE HEARTED
152.	SOUR	
153.	STABLE EMOTIONALLY	CHANGEABLE
154.	STRONG IN PERSONALITY	
155.	SUBJECTIVE	GUIDED BY REALITY
156.	SUGGESTIBLE	
157.	TACTFUL	TACTLESS
158.	TALKATIVE	TACITURN
159.	TEMPERATE	EXTREME (in schizothyme sense)
160.	THOUGHTFUL	UNREFLECTIVE
161.	THRIFTY	CARELESS WITH GOODS
162.	TIMID (Disposition)	ADVENTUROUS
163.	TREACHEROUS	
164.	TRUSTFUL	SUSPICIOUS
165.	VERSATILE	
166.	VINDICTIVE	UNRESENTFUL
167.	VIVACIOUS	
168.	WANDERING	SETTLING DOWN
169.	WITTY	HUMORLESS
170.	WISE	FOOLISH
171.	WORRYING	PLACID

Fonte: CATTELL (1943b), pp. 493-496.

Mais ainda, a partir destes traços, Catell proporcionou uma redução ainda menor de traços por meio de análise de clusters obtidos com coeficientes e correlação nas respostas dos questionários.

Por meio desta técnica, conseguiu agrupar esse conjunto de traços em *clusters*. Estes conjuntos obtidos apresentam graus variados de sobreposição (interseção) entre si. A partir desta clusterização dos traços, o autor procedeu a uma redução maior dos mesmos, adotando as seguintes medidas:

Os traços que obtiveram correlação nas respostas superior a 0,8 foram considerados idênticos na descrição do comportamento a que se referiam. As justificativas para tal, nas palavras do autor foram (CATTELL, 1943b, p.497):

Since a separate study had shown that the judges had a reliability of about 0.7 to 0.8 in estimating these traits, any traits correlating above 0.8 were considered identical, i.e., as instances of our having cautiously split a trait element further than was necessary.

Traços com grau de importância prática razoável, que não mostraram correlações significantes com outros traços foram considerados como sendo variáveis independentes das demais.

Clusters sem interseção com os demais, consistindo de quatro ou mais traços foram representados por uma variável, representando todo o comportamento comum àquele conjunto.

Com estas medidas, reduziu o conjunto de *clusters* iniciais para 60 *clusters*, contendo 135 traços. A seguir, o autor conseguiu reduzir (CATTELL, 1945) ainda mais o número de *clusters* por:

Considerando como um cluster único *clusters* remanescentes com grande parcela de seus traços constituintes comuns uns aos outros.

Ignorando *clusters* menores, menos confiáveis.

Selecionando, dos *clusters* remanescentes, somente aqueles já confirmados por outros pesquisadores.

Estas medidas propiciaram a seleção final de 35 *clusters*, contendo 105 traços. Cada *cluster* passa a representar uma variável de entrada para a análise fatorial que o autor aplicará a seguir.

Assim, as 35 variáveis selecionadas para aplicações envolvendo análise fatorial e seus respectivos traços componentes são apresentados na tabela 13 abaixo: as 35 variáveis selecionadas, por meio de análise de *clusters* por CATTELL (1943b e 1945) para aplicações com análise fatorial.

Tabela 13: As 35 variáveis adotadas por Cattell no modelo de 12 fatores da personalidade (12PF).

POSITIVE POLE	NEGATIVE POLE
1. <i>Self Assertive</i>	<i>Self Submissive</i>
Boastful	Modest
Assertive	Submissive
Conceited	Self critical, dissatisfied
2. <i>Intelligent, analytical</i>	<i>Unimaginative, stupid</i>
Intelligent	Stupid
Clear thinking	Incoherent, confused
Clever	

3. <i>Wise, mature, polished</i> Independent Reliable Mature	<i>Dependent, silly, incoherent</i> Emotionally dependent Undependable Emotionally immature, irresponsible
4. <i>Changeable, frivolous</i> Unreflective Impulsive Profligate	<i>Thoughtful, stoic, reserved</i> Thoughtful Deliberate Austere
5. <i>Neurotic</i> Self-deceiving Hypochondriacal Nervous, specific neurotic symptoms	<i>Not generally neurotic</i> Realistic
6. <i>Hard, cynical</i> Thankless Hostile Hardhearted	<i>Kindly, gentle, idealistic</i> Grateful Friendly, understanding Softhearted
7. <i>Wilful, egotistic, predatory</i> Extra-punitive (blaming mistakes on others) Headstrong Exhibitionist	<i>Mild, self-effacing, tolerant</i> Gentle tempered Self-effacing <i>Adaptable, friendly</i>
8. <i>Rigid, tyrannical, vindictive</i> Extra-punitive Inflexible (emotionally) Hostile	 Adaptable (to change) Friendly
9. <i>Surly, hard</i> Thankless Hardhearted, embittered Short-tempered	<i>Good-natured, easy-going</i> Grateful Softhearted Easy-going
10. <i>Demoralized, autistic</i> Unrealistic Quitting Subjective, evasive	<i>Realistic, facing life</i> Realistic, practical Persevering Facing life
11. <i>Strong-willed, conscientious</i> Persevering Painstaking Conscientious	<i>Indolent, incoherent, impulsive</i> Quitting Slipshod Conscienceless
12. <i>Intellectual</i> Thoughtful Analytical Wide interests	<i>Simple, undisciplined mind</i> Unreflective Narrow interests
13. <i>Insecure, infantile, hostile</i> Easily jealous Thankless, unappreciative Self-pitying	<i>Mature, kind, tactful</i> Grateful
14. <i>Anti-social, schizoid</i> Cynical Obstructive Timid, withdrawn	<i>Out-going, idealistic, cooperative</i> Idealistic Cooperative Adventurous

POSITIVE POLE	NEGATIVE POLE
15. <i>Cheerful, enthusiastic, witty</i> Genial Optimistic Enthusiastic	<i>Unhappy, frustrated, dour</i> Cold-hearted Pessimistic Apathetic
16. <i>Active, neurotic, creatively unstable</i> Intrusive Impulsive Neurotic	<i>Self-controlled, rigid, conventional</i> Reserved Deliberate
17. <i>Character neurosis, psychopathic</i> (<i>Neurotic, irritable, uncontrolled</i>) Fickle Dishonest	<i>Emotionally mature</i> Balanced Loyal Honest
18. <i>High-strung, expressive, driven</i> Highly strung Hurried Vivacious	<i>Phlegmatic</i> (<i>Unexcited and unexcitable</i>) Lethargic Lethargic
19. <i>Spiteful, tight-fisted, superstitious</i> Hostile Obstructive Secretive (<i>Irrational, obsessive fears</i>)	<i>Natural, friendly, open</i> Friendly Cooperative Frank
20. <i>General emotionality (with maladjustment)</i> Emotional (in all ways) Dissatisfied Excitable	<i>Unemotional</i> Unemotional Content Phlegmatic
21. <i>Ascendant, expressive, widely interested</i> Energetic, spirited Self confident Debonnaire	<i>Retiring, quiet, narrow</i> Languid Self distrustful
22. <i>Responsive, genial, sentimental</i> Responsive Genial Social interests	<i>Aloof, cold, misanthropic</i> Aloof Coldhearted Brooding (not an opposite)
23. <i>Facile, foppish, affected</i> Exhibitionist Eloquent Flattering	<i>Inarticulate, natural</i> Self effacing Inarticulate Natural (not an opposite)
24. <i>Hostile, paranoid</i> Sadistic Suspicious Mulish	<i>Trustful, good-tempered</i> Not sadistic Trustful Reasonable
25. <i>Aesthetic interests, independent mind</i> General aesthetic interests Musical ability and interests Independent	
26. <i>Restlessly, sthenically, hypomanically emotional</i> Emotional Excitable Impatient (<i>Sthenically emotional—[Burt]</i>)	<i>Calm, self-effacing, patient</i> Unemotional Phlegmatic Patient

POSITIVE POLE	NEGATIVE POLE
27. <i>Infantile, demanding, self-centered</i> Infantile Self-pitying Exhibitionist	<i>Emotionally mature, adjusting to frustration</i> Mature emotionally Self effacing
28. <i>Changeable, characterless, unrealistic</i> Changeable	<i>Stable, integrated character</i> Stable emotionally Self-respecting Self-controlled
29. <i>Psychophysically vigorous, alert</i> Alert Energetic, spirited Quick	<i>Neuroathenic</i> Absent-minded Languid Slow
30. <i>Adventurous, lusty</i> Incontinent Gluttonous Curious	<i>Generally inhibited, timid</i> Inhibited Queasy Unenquiring
31. <i>Sociable, hearty</i> Sociable (forward, gregarious) Responsive Hearty	<i>Seclusive, shy</i> Shy (and seclusive) Aloof Quiet
32. _____ Optimistic Placid	<i>Melancholic (agitated/involuntal)</i> Hypochondriacal Pessimistic Worrying
33. <i>Tough, solid, talkative</i> Tough Lethargic	<i>Introspective, sensitive, scared</i> Introspective Sensitive Hurried
34. <i>Imaginative, introspective, constructive</i> Labile Intuitive Careless of material things	<i>Set, smug, thrifty</i> Habit-bound Logical (precise) Thrifty
35. <i>Smart, assertive</i> Sophisticated Intelligent Assertive	<i>Simple-hearted, meek</i> Simple Stupid Submissive

Fonte: CATTELL (1945b, pp. 71-74).

Muitas dessas 35 variáveis também estão presentes ao longo dos cinco fatores de personalidade encontrados em estudos subsequentes - especialmente em TUPES e CRISTAL (1961), PEABODY e GOLDBERG (1989) e MACRAE e OLIVER (1992). Este fato também pode ser verificado na tabela 17 abaixo, onde se verifica que muitas dessas 35 variáveis também estão presentes em outros modelos de personalidade construídos por outros autores.

4.2.2.2.2 O desenvolvimento do Questionário dos 16 Fatores da Personalidade

Tendo chegado às mencionadas 35 variáveis (ou 35 *clusters* finais) o autor então procede à utilização dos mesmos em experimentos com indivíduos usando análise fatorial (como se verá tal experimento gerará um modelo de 12 fatores de personalidade que servirá de insumo para o posterior desenvolvimento do modelo de 16 fatores).

Tais *clusters* correspondem ao que o autor chamava de traços de superfície (*surface traits*), que se diferenciam dos traços de origem (*source traits*). Ao recorrer ao emprego de análise fatorial com experimentos com indivíduos, o autor passa então a buscar os *source traits* (correspondentes aos fatores na análise fatorial). De acordo com o autor:

The factors and clusters corresponding to source and surface traits have to be derived from observations both on intra-individual and inter-individual covariation of measurements and are basically of two kinds, (a) environmental mold traits, in which the factor lies in (or was originally in) the environment, and constitutional traits in which the unity of covariation springs from some factor of dynamic, temperamental or ability character, or of all three together, as they are resident within the genetic pattern of the organism.. CATTELL (1945b, p. 69).

O autor ainda continua, com a importante explicação de que:

The second article began the experimental applications of the methods implied by these principles. It introduced also the notion of the 'personality sphere, maintaining that a true perspective in the factor analysis of such a new realm as that of personality can only be obtained by beginning with an essentially complete trait-variable population-namely, that representing the entire 'surface' of the personality sphere. The use of the personality sphere necessitates basing the factor analysis on trait variables which are verbally defined, compelling one to commence with behavior ratings rather than with objective test-measurements. The argument of the introductory article 3 is that the entire personality sphere, i.e. the entire range of personality aspects, is only to be found in the accumulated symbolism of language which alone has long mirrored all the facets of human nature that are important to man. By processes described in detail elsewhere, 4 a truly representative list of personality traits was derived from language; a basic list which itself may have some permanent value as a set of landmarks for researches in this field. We must regard the use of verbally-defined, rated traits, however, as only a temporary phase in personality research - as a preliminary scaffolding or framework for giving perspective, and within which objective, miniature situations, 5 projections, 6. or other precise, objective test-measurements can later be introduced to carry the weight of evidence. CATTELL (1945b, pp. 69-70).

Por meio da aplicação da análise fatorial com as 35 variáveis em uma população de 208 homens adultos de diversas ocupações, Cattell identificou 12 fatores capazes de explicar consistentemente bem a personalidade. Estes fatores estão listados na tabela abaixo:

Tabela 14: O modelo de 12 fatores de personalidade de Cattell.

LIST OF CONTINGENT DESCRIPTIONS OF FACTORS	
Factor	
A	Cyclothyme-v-Paranoid Schizothyme
B	General Mental Capacity (Spearman's 'g,' in personality expression)
C	Emotionally Mature, Stable Character-v-General Emotionality
D	Hypomanic, Sthenic Emotionality-v-Phlegmatic Frustration—Tolerance
E	Dominance-v-Submissiveness
F	Surgency-v-Melancholy, Shy, Desurgency
G	Positive Character Integration-v-Immature, Dependent Character
H	Charitable, Adventurous Surgency-v-Inhibited, Insecure Desurgency
I	Sensitive, Imaginative, Neurotic Emotionality-v-Rigid, Tough Poise
J	Neurasthenic-v-Vigorous "Obsessional" Character
K	Trained, Cultured Mind-v-Boorishness
L	Rhathymic, Adjusted, Surgency-v-Schizoid Desurgency

Fonte: CATTELL (1945b, p. 89).

Em novo experimento (CATTELL, 1947), o autor confirma os 12 fatores de personalidade como amplamente capazes de explicar os comportamentos humanos.

Aplicações subsequentes por este autor identificaram que 16 fatores seriam necessários para a completa descrição da personalidade humana. Este modelo de 16 fatores gerou diversos questionários, que passaram a ser identificados como 16 PF *questionnaires*.

Uma versão simplificada deste questionário é apresentada por Cattell em seu trabalho *A Shortened Basic English Version of the 16PF Questionnaire* (CATTELL, 1956).

No que concerne aos 16 fatores da personalidade identificados posteriormente por Cattell, a tabela a seguir apresenta um detalhamento destes fatores, com os termos descritores de cada um deles.

Tabela 15: Os 16 Fatores da personalidade identificados por Raymond Cattell

<i>Descriptors of Low Range</i>	<i>Primary Scales</i>	<i>Descriptors of High Range</i>
<i>Reserved, Impersonal, Distant</i>	<i>Warmth (A)</i>	<i>Warm-hearted, Caring, Attentive To Others</i>
<i>Concrete, Lower Mental Capacity</i>	<i>Reasoning (B)</i>	<i>Abstract, Bright, Fast- Learner</i>
<i>Reactive, Affected By Feelings</i>	<i>Emotional Stability (C)</i>	<i>Emotionally Stable, Adaptive, Mature</i>
<i>Deferential, Cooperative, Avoids Conflict</i>	<i>Dominance (E)</i>	<i>Dominant, Forceful, Assertive</i>
<i>Serious, Restrained, Careful</i>	<i>Liveliness (F)</i>	<i>Enthusiastic, Animated, Spontaneous</i>
<i>Expedient, Nonconforming</i>	<i>Rule-Consciousness (G)</i>	<i>Rule-Conscious, Dutiful Socially Bold, Venturesome,</i>
<i>Shy, Timid, Threat-Sensitive</i>	<i>Social Boldness (H)</i>	<i>Thick-Skinned Sensitive, Aesthetic, Tender- Minded</i>
<i>Tough, Objective, Unsentimental</i>	<i>Sensitivity (I)</i>	<i>Vigilant, Suspicious, Skeptical, Wary</i>
<i>Trusting, Unsuspecting, Accepting</i>	<i>Vigilance (L)</i>	<i>Abstracted, Imaginative, Idea-Oriented</i>
<i>Practical, Grounded, Down-To- Earth</i>	<i>Abstractedness (M)</i>	<i>Private, Discreet, Non- Disclosing</i>
<i>Forthright, Genuine, Artless</i>	<i>Privateness (N)</i>	<i>Apprehensive, Self- Doubting, Worried</i>
<i>Self-Assured, Unworried, Complacent</i>	<i>Apprehension (O)</i>	<i>Open To Change, Experimenting</i>
<i>Traditional, Attached To Familiar</i>	<i>Openness to Change (Q1)</i>	<i>Self-Reliant, Solitary, Individualistic</i>
<i>Group-Orientated, Affiliative</i>	<i>Self-Reliance (Q2)</i>	<i>Perfectionistic, Organized, Self-Disciplined</i>
<i>Tolerates Disorder, Unexacting, Flexible</i>	<i>Perfectionism (Q3)</i>	
<i>Relaxed, Placid, Patient</i>	<i>Tension (Q4)</i>	<i>Tense, High Energy, Driven</i>

Fonte: CATTELL E MEAD (2007).

Cattell posteriormente identificou uma associação entre estes fatores com as cinco, mais amplas grandes dimensões da personalidade (CATTELL e MEAD, 2007). A figura a seguir ilustra esta correspondência.

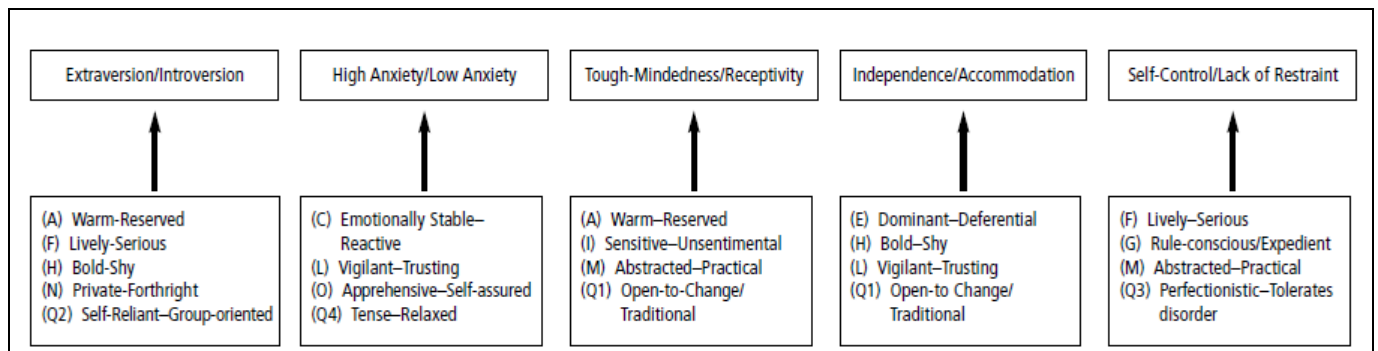


Figura 22: Correspondência entre os 16 fatores identificados por Cattell e as cinco dimensões amplas da personalidade de Cattell.

Fonte: CATTELL E MEAD, 2007, p. 138.

4.2.2.2.3 Estudo de Norman (1963)

Norman (1963) utilizou um subconjunto de 20 variáveis das 35 propostas inicialmente por Cattell. Agrupou essas 20 variáveis descritivas de personalidade em 5 grandes fatores: Extraversão (*Extraversion*), Agradabilidade (*Agreeableness*), Consciência (*Conscientiousness*), Estabilidade Emocional (*Emotional Stability*) e Cultura (*Culture*).

Com este agrupamento, batizou este resultado de Os Cinco Grandes Fatores da Personalidade (ou Modelo dos Cinco fatores, *Five Factor Model*).

A tabela a seguir mostra as 20 variáveis agrupadas nos cinco grandes fatores.

Tabela 16: Os Cinco Fatores de Norman

<i>Extraversion</i>	1	talkative	- silent
	2	frank, open	- secretive
	3	adventurous	- cautious
	4	sociable	- reclusive
<i>Agreeableness</i>	5	good-natured	- irritable
	6	not jealous	- jealous
	7	mild, gentle	- headstrong
	8	cooperative	- negativistic
<i>Conscientiousness</i>	9	fussy, tidy	- careless
	10	responsible	- undependable
	11	scrupulous	- unscrupulous
	12	persevering	- quitting, fickle
<i>Emotional Stability</i>	13	poised	- nervous, tense
	14	calm	- anxious
	15	composed	- excitable
	16	not hypochondriacal	- hypochondriacal
<i>Culture</i>	17	artistically sensitive	- insensitive
	18	intellectual	- unreflective, narrow
	19	polished, refined	- crude, boorish
	20	imaginative	- simple, direct

FONTE: DE RAAD, 2000, p.3.

Esses 171 termos foram posteriormente condensados em 35 variáveis utilizando uma redução correlacional baseada em gradações (*ratings*) de 100 adultos, obtidos por meio da aplicação de questionários aos mesmos.

As 35 variáveis podem ser vistas na tabela a seguir, que ilustra também, como estas mesmas 171 palavras foram agrupadas em variáveis em outros estudos posteriores ao de Cattell.

Tabela 17: Presença comparada das 171 variáveis selecionadas por Cattell em diversos modelos de fatores de personalidade.

<i>Name of Bipolar Trait</i>		<i>Norman</i>	<i>T&C*</i>	<i>Fiske</i>
1	Ready to cooperate vs Obstructive	8	1	1
2	Emotionally stable vs Changeable		2	2
3	Attention-getting vs Self-sufficient			15
4	Assertive, self-assured vs Submissive		3	3
5	Depressed, solemn vs Cheerful			4
6	Frivolous vs Responsible	10	4	5
7	Attentive to people vs Cool, aloof		5	6
8	Easily upset vs Unshakable, poised, tough	13	6	7
9	Languid, slow vs Energetic, alert		7	
10	Boorish vs Intellectual, cultured	18	8	8
11	Suspicious vs Trustful		9	9
12	Good-natured, easygoing vs Spiteful, grasping, critical	5	10	10
13	Calm, phlegmatic vs Emotional	14	11	22
14	Hypochondriacal vs Not so	16	12	
15	Mild, self-effacing vs Self-willed, egotistic	7	13	
16	Silent, introspective vs Talkative	1	14	11
17	Persevering, determined vs Quitting, fickle	12	15	
18	Cautious, retiring, timid vs Bold, adventurous	3	16	12
19	Hard, stern, vs Kindly, soft-hearted		17	
20	Insistently orderly vs Relaxed, indolent	9	18	
21	Polished vs Clumsy, awkward	19	19	13
22	Prone to jealousy vs Not prone to jealousy	6	20	
23	Rigid vs Adaptable		21	14
24	Demanding, impatient vs Emotionally mature		22	
25	Unconventional, eccentric vs Conventional		23	
26	Placid vs Worrying, anxious		24	16
27	Conscientious vs Somewhat unscrupulous	11	25	17
28	Composed vs Shy, bashful	15		
29	Sensitively imaginative vs Practical, logical	20		18
30	Neurotic fatigue vs Absence of neurotic fatigue		26	
31	Esthetically fastidious vs Lacking artistic feeling	17	27	
32	Marked interest in opp. sex vs Slight interest in opp. sex			19
33	Frank, expressive vs Secretive, reserved	2	28	20
34	Gregarious, sociable vs Self-contained	4	29	
35	Dependent, immature vs Independent-minded		30	21

*) T&C is Tupes and Christal (1961)

Fonte: DE RAAD, 2000, p. 6. As 35 variáveis (traços) atribuídas a Raymond Cattell foram as constantes do seu modelo de 12 fatores da personalidade, confirmado empiricamente em seu trabalho *Confirmation and Classification of Primary Personality Factors* (1947). Cada uma dessas 35 variáveis corresponde a um cluster obtido em trabalho anterior (1943b). Nesta mesma tabela se apresenta as variáveis (fatores) presentes em outros modelos da personalidade desenvolvidos por TUPES e CRISTAL (1961) e FISKE (1949), onde se pode constatar que muitos traços são comuns aos três modelos. As numerações dos traços apresentadas acima são as numerações que os mesmos recebem, pelos autores citados, nos três modelos

de personalidade – traços que não recebem numeração em uma determinada coluna, na tabela acima, não estão presentes no modelo correspondente a esta coluna-.

4.2.2.2.4 Estudo de Tupes e Cristal (1961)

Ernet Tupes e Raymond Christal (1961) encontraram 5 fatores recorrentes a partir do questionário de 16 fatores da personalidade de Cattell. Estes fatores recorrentes foram chamados por estes autores de *Surgency*, *Agreeableness*, *Dependability*, *Emotional Stability* e *Culture*. Nesses cinco fatores o modelo encontrado por estes autores contava com 30 variáveis (traços), como pode ser constatado na tabela acima.

A tabela a seguir apresenta o modelo de cinco de fatores de TUPES e CRISTAL, e as variáveis-traço pertencentes a cada fator.

Tabela 18: Modelo dos Cinco Grandes Fatores de Tupes e Cristal.

<u>Trait Variable</u>	
<u>No.</u>	<u>Name</u>
14	Silent vs Talkative
28	Secretive vs Frank
16	Cautious vs Adventurous
3	Submissive vs Assertive
29	Self-contained vs Sociable
7	Languid, Slow vs Energetic
10	Spiteful vs Goodnatured
20	Jealous vs Not So
22	Demanding vs Emotionally Mature
13	Self-willed vs Mild
1	Obstructive vs Cooperative
9	Suspicious vs Trustful
21	Rigid vs Adaptable
17	Hard, Stern vs Kindly
5	Cool, Aloof vs Attentive to People
18	Relaxed, Indolent vs Insistently Orderly
4	Frivolous vs Responsible
25	Unscrupulous vs Conscientious
15	Quitting vs Persevering
23	Unconventional vs Conventional
26	Neurotic vs Not So
24	Worrying, Anxious vs Placid
6	Easily Upset vs Poised, Tough
12	Hypochondriacal vs Not So
11	Emotional vs Calm
2	Changeable vs Emotionally Stable
8	Boorish vs Intellectual, Cultured
34	Practical, Logical vs Imaginative
19	Clumsy, Awkward vs Polished
30	Immature vs Independent-Minded

Fonte: TUPES e CRISTAL (1961), p. 13.

Deste modo pode-se verificar que o primeiro fator do modelo destes autores contém os traços *silent*, *secretive*, *cautious*, *submissive*, *self-contained* e *languid* (conjuntamente com os traços representando as ideias opostas a esses conceitos).

O segundo fator contém os traços *spiteful*, *jealous*, *demanding*, *self-willed*, *obstructive*, *suspicious*, *rigid*, *hard* e *cool* (e seus conceitos opostos).

O terceiro fator contém *relaxed*, *frivolous*, *unscrupulous*, *quitting* e *unconventional* (conjuntamente com seus conceitos opostos).

O quarto fator contém *neurotic, worrying, easly upset, hypochondriacal, emotional e changeable* (e seus conceitos opostos).

O último fator contém os traços *boorish, practical, clumsy e imature*.

4.2.2.3 Os modelos de cinco fatores de personalidade

Como visto anteriormente, diversos estudos independentes chegaram a modelos de personalidade compostos por cinco fatores. Anteriormente se apresentou os modelos de O modelo dos cinco fatores globais de Raymond Cattell (CATTELL e MEAD, 2007), o modelo de cinco fatores de TUPES e CRISTAL (1961), o modelo de NORMAN (1963) e o modelo de MACRAE e OLIVER (1992).

A construção destes modelos foi fruto de importantes contribuições, efetuadas por diversos pesquisadores de modo independente, ao longo de diversas décadas. Diversos grupos (a seguir apresentados) de pesquisadores trabalharam de modo independente na área de traços de personalidade e de uma forma geral identificaram os mesmos cinco fatores explicativos da personalidade (embora estes os fatores identificados por estes grupos tivessem diferentes nomes).

Existe um razoável consenso na literatura de que os principais autores que contribuíram para o desenvolvimento deste modelo foram: CATTELL (1943a, 1943b, 1944, 1945, 1946a, 1946b, 1946c, 1956 e 1957) da Universidade de Illinois, TUPES e CRISTAL (1961), NORMAN (1963), NORMAN e GOLDBERG (1966), GOLDBERG (1992, 1993) do *Oregon Research Institute*, PEABODY e GOLDBERG (1989), SAUCIER e GOLDBERG (1996), COSTA E MACCRAE (1976, 1985, 1987, 1992 e 2000) do *National Institutes of Health* e MACCRAE E OLIVER (1992).

Do modelo de cinco fatores de TUPES e CRISTAL (1961), DIGMAN (1989) o aperfeiçoou. Este modelo chegou a sua forma final (e atualmente aceita) por GOLDBERG (1993) esforço de construir tal modelo,

A tabela a seguir apresenta o modelo de Cinco Fatores da Personalidade de acordo com MACRAE e OLIVER (1992), e os traços pertencentes a cada um deles.

Tabela 19: Modelo dos Cinco Grandes Fatores da Personalidade de acordo com MACRAE e OLIVER (1992).

Factor		Factor definers		
Name	Number	Adjectives ^a	Q-sort items ^b	Scales ^c
Extraversion (E)	I	Active	Talkative	Warmth
		Assertive	Skilled in play, humor	Gregariousness
		Energetic	Rapid personal tempo	Assertiveness
		Enthusiastic	Facially, gesturally expressive	Activity
		Outgoing	Behaves assertively	Excitement Seeking
		Talkative	Gregarious	Positive Emotions
Agreeableness (A)	II	Appreciative	Not critical, skeptical	Trust
		Forgiving	Behaves in giving way	Straightforwardness
		Generous	Sympathetic, considerate	Altruism
		Kind	Arouses liking	Compliance
		Sympathetic	Warm, compassionate	Modesty
		Trusting	Basically trustful	Tender-Mindedness
Conscientiousness (C)	III	Efficient	Dependable, responsible	Competence
		Organized	Productive	Order
		Planful	Able to delay gratification	Dutifulness
		Reliable	Not self-indulgent	Achievement Striving
		Responsible	Behaves ethically	Self-Discipline
		Thorough	Has high aspiration level	Deliberation
Neuroticism (N)	-IV	Anxious	Thin-skinned	Anxiety
		Self-pitying	Brittle ego defenses	Hostility
		Tense	Self-defeating	Depression
		Touchy	Basically anxious	Self-Consciousness
		Unstable	Concerned with adequacy	Impulsiveness
		Worrying	Fluctuating moods	Vulnerability
Openness (O)	V	Artistic	Wide range of interests	Fantasy
		Curious	Introspective	Aesthetics
		Imaginative	Unusual thought processes	Feelings
		Insightful	Values intellectual matters	Actions
		Original	Judges in unconventional terms	Ideas
		Wide interests	Aesthetically reactive	Values

Fonte: MACRAE e OLIVER, 1992, pp. 178-179.

Os cinco fatores acima apresentados são interpretados da seguinte forma:

- *Abertura (Openness)*: representa aspectos da personalidade caracterizados pela abertura a novas experiências, pela apreciação por artes, emoções, aventuras, ideias e crenças não usuais. Pessoas que são abertas às experiências são intelectualmente curiosas e sensíveis, abertas a emoção, interessadas em arte e em se tentar coisas novas (experiências novas).

De acordo com o modelo de Cinco Fatores de MACRAE e OLIVER (1992) os traços contidos neste fator são: *active, assertive, energetic, enthusiastic, outgoing e talkative*.

- *Conscientiousness (Consciência)*: esta dimensão da personalidade está relacionada à tendência de se mostrar disciplina e autodisciplina, de se agir obedientemente e do foco em se sair bem sucedido em relação à testes e expectativas. No modelo de Cinco Fatores de

MACRAE e OLIVER (1992) os traços contidos neste fator são: *efficient, organized, planful, reliable, responsible* e *thorough*. Altos *scores* em *conscientiousness* indicam comportamentos com predominância de ações planejadas (ao invés de ações espontâneas ou impulsivas).

-*Extraversion* (extroversão): esta dimensão da personalidade está relacionada com o grau de engajamento social e energia do indivíduo. Indivíduos muito extrovertidos apreciam interagir com pessoas, e são geralmente percebidos por terceiros como sendo pessoas energéticas. Na figura acima se pode verificar que os traços (ou variáveis) pertencentes a este fator são: *active, assertive, energetic, enthusiastic, outgoing* e *talkative*.

- *Agreeableness* (socialização): este fator reflete as diferenças individuais no que se refere à preocupação com a harmonia social. Indivíduos que são avaliados com scores muito elevados neste fator tendem a serem gentis, generosos, confiáveis, e valorizam estar e se relacionar com outros indivíduos.

No modelo apresentado acima, os traços *appreciative, forgiving, generous, kind, sympathetic* e *trusting* pertencem a este fator.

-*Neuroticism* (neuroticismo): este fator está relacionado com comportamentos que revelam tendências a se experimentar/vivenciar emoções ruins, tais como raiva, ansiedade e depressão. Aqueles indivíduos que apresentam altos *scores* neste fator são emocionalmente reativos e vulneráveis ao estresse. Nesses indivíduos, as reações emocionais negativas tendem a persistir por períodos mais longos.

4.2.2.4 Equivalências entre os diferentes modelos de cinco fatores da personalidade

Como apresenta CATTELL e MEAD (2007), estudos mais recentes tem verificado uma correspondência entre diferentes modelos de cinco fatores, baseando-se em técnicas de análise fatorial, mais modernas.

A tabela a seguir apresenta o alinhamento entre três dos principais modelos de cinco fatores na literatura.

Tabela 20: Alinhamento entre três dos principais modelos de cinco fatores de personalidade

<i>16PF (Cattell)</i>	<i>NEO-PI-R (Costa and McCrae)</i>	<i>Big Five (Goldberg)</i>
Extraversion/Introversion	Extraversion	Surgency
Low Anxiety/High Anxiety	Neuroticism	Emotional stability
Tough-Mindedness/Receptivity	Openness	Intellect or culture
Independence/Accommodation	Agreeableness	Agreeableness
Self-Control/Lack of Restraint	Conscientiousness	Conscientiousness or dependability

Fonte: Cattell e Mead, 2007, p. 141.

4.2.2.5 Influência de fatores genéticos

Diversos estudos tem revelado a importância de variáveis genéticas na personalidade humana. Estes estudos revelaram não só a importância de variáveis desse tipo para a explicação da personalidade humana como a compatibilidade dessas influências com o modelo dos Cinco Grandes Fatores.

Exemplos desses estudos são o realizado por JANG e LIVESLEY (1996) que ao realizarem experimentos com gêmeos, sugeriram que tanto a hereditabilidade como fatores ambientais contribuem para os traços nos cinco fatores. Estes autores verificaram que as influências de fatores genéticos nas cinco dimensões de Neuroticismo, Extroversão, Abertura, Socialização e Consciência foram estimadas, respectivamente, em 41%, 53%, 61%, 41% e 44%, usando como amostra 123 pares de gêmeos idênticos e 127 pares de gêmeos fraternos.

BOUCHARD (2004), também estimou os efeitos da hereditabilidade nos cinco fatores analisados, chegando a conclusão que, esses efeitos, sobre os fatores Neuroticismo, Extroversão, Abertura, Socialização e Consciência foram estimados, respectivamente, em 48%, 54%, 57%, 42% e 49%.

4.2.2.6 Influência das diferenças culturais sobre os cinco fatores

Alguns estudos têm revelado certas dificuldades do modelo dos Cinco Fatores da Personalidade na explicação de comportamentos quando se varia a cultura de povos.

O estudo de CHEUNG, F., M. et al (2011) sugere que o fator Abertura não tem suporte em países asiáticos e que um novo quinto fator precisa ser definido nestas regiões.

Segundo estes autores, a evidência aponta para o fato de que este fator é mais relevante para a cultura ocidental.

Outros estudos encontraram mais do que cinco fatores após aplicarem análise fatorial a um conjunto de dados. Tal é o caso do estudo de SINGH e MISRA (2010), utilizaram 357

traços escolhidos de um *thesaurus* específico na construção de um questionário com escalas de três pontos para cada um desses traços, aplicados em indianos na língua hindu.

A partir das correlações nas respostas aplicaram análise fatorial e identificaram seis fatores, a saber:

- *Rajasic*: fator que representa a tendência a ser amigável e generoso.

- *Sattvic*: característicos dos indivíduos que possuem qualidades sociais e espirituais.

- *Tamasic*: compreendendo características como ser não civilizado, sem consideração, arrogante, ciumento e intolerante.

- *Competence*: compreendendo as características de ser autoconfiante e de ser dinâmico e buscar o progresso.

- *Neuroticism*: compreendendo características como instabilidade emocional e ansiedade.

- *Extraversion*: contendo características como ser assertivo, extrovertido, sem medo, corajoso e diplomático.

JACKSON e PAUNONEN (2000) também verificaram dificuldades do modelo *Big Five* com variações de cultura.

4.2.2.7 Críticas aos modelos de cinco fatores da personalidade

As principais críticas aos modelos de cinco fatores de personalidade repousam na dificuldade dos cinco fatores de qualquer um desses modelos de sistematicamente conseguir explicar as diferenças individuais nos indivíduos ao longo de diversas culturas.

Como visto modelos compostos por cinco fatores tem enfrentado dificuldades de explicar estas diferenças, sobretudo nas sociedades orientais (a esse respeito vide JACKSON e PAUNONEN, 2000 e SINGH e MISRA, 2010), de modo que a pesquisa na área ainda enfrenta adversidades desse tipo.

5 SISTEMAS DESENVOLVIDOS, BANCOS DE DADOS, ROTINAS DESENVOLVIDAS, PROGRAMAS UTILIZADOS E CARACTERÍSTICAS DO *THESAURUS* ADOTADO

Neste capítulo se procederá à apresentação dos códigos, bancos de dados e do sistema desenvolvido neste trabalho com fins de subsidiar seus desenvolvimentos.

Porém, antes de se proceder à apresentação deste sistema, *softwares*, linguagens de programação e bancos de dados utilizados, se irão apresentar as características do *thesaurus* adotado.

O banco de palavras adotado, em inglês, é produzido por TT-*Software/Databases* e disponível para compra nos endereços eletrônicos www.language-databases.com e www.true-term-mobile.com.

Este banco está disponível em diversos formatos eletrônicos: em formato de banco de dados compatível com o Microsoft Access, um formato em planilha para o *software* Microsoft Excel e um último formato para banco de dados em SQL.

Como em todo *thesaurus*, este repositório de palavras oferece uma listagem de palavras similares ou relacionadas, porém não necessariamente sinônimas.

Utilizou-se, para os propósitos desta tese o banco em formato de planilha Excel, pois: primeiramente a visualização e algumas manipulações simples do banco (como trocar as linhas do banco de ordem) são possíveis e não demandam a construção de nenhum código específico para tal. Em segundo lugar, não se conseguiu importar o banco em formato SQL no banco Postgres SQL utilizado no primeiro sistema desenvolvido (em linguagem php) para tratamento das palavras e produção de relatórios de interesse para o trabalho.

Este último fator foi decisivo para a escolha deste formato de *thesaurus* a ser importado pelos sistemas desenvolvidos.

Finalmente, se escolheu empregar a versão no idioma em inglês deste banco de palavras, pois boa parte dos estudos de *clusters* de palavras em *thesauri* deste tipo são feitos neste idioma, de forma que alguns resultados obtidos neste trabalho podem ser mais facilmente comparados a resultados que possam ser obtidos, empregando-se as metodologias aqui adotadas, em outros *thesauri*, como por exemplo, o Wordnet.

5.1 ESTRUTURA DO BANCO EM PLANILHA EXCEL

A versão do *thesaurus* em planilha Excel está estruturada da seguinte forma exemplificada por meio de uma imagem de um pedaço banco abaixo:

1	Keyword	Bed 1	Bed 2	Bed 3	Bed 4
2	intelligent	bright [a]: sharp; quick; intellectual; unders	keen [a]: ast	rational [a]: logical; reaso	
3	pictorial	graphic [a]: representative; illustrative; det;	illustrated [a]: pictured; painted; embe		
4	exact	characteristic [a]: demanding; exacting; rigi	characteristi	property [a]: means [v]: co	
5	deft	proficient [a]: adept; skilled; accomplished; master; skilful; expert; practiced			
6	dexterous	adroit [a]: deft; clean; adept			
7	proficient	qualified [a]: expert; competent; accomplished; capable; fitted; suited			
8	skilled	facile [a]: proficient; practiced; skilful; accomplished; expedite; aid; adept			

Figura23: Esquema de estruturação do *thesuarus* em inglês

FONTE: *thesaurus* da TT Software/Databases.

A figura acima apresenta como se dá a associação entre palavras no *thesaurus* utilizado, no formato de planilha Excel. Na primeira coluna da planilha têm-se as palavras-chave (*keywords*) a partir das quais o banco de palavras irá relacionar outras palavras.

Ao longo de uma determinada linha qualquer, têm-se, para uma determinada palavra-chave as palavras associadas a ela ao longo de células nesta linha. As palavras dentro de uma célula, relacionadas a esta palavra-chave, se relacionam a esta última segundo um dado sentido de associação. Este sentido de associação está expresso junto com as palavras que se relacionam à *keyword*, na mesma célula em que se encontra, só que antes da letra dentro do colchete.

Finalmente, esta letra dentro do colchete indica a classe gramatical cumprida por estas palavras dentro desta célula, na associação destas com a palavra-chave.

Assim, pegando-se, por exemplo, a palavra *intellectual*, no exemplo dado na tabela acima, vê-se que ela está associada com a palavra-chave *intelligent*, no sentido de *bright*, cumprindo assim a função de adjetivo.

Segundo os linguistas que prepararam este banco, as classes gramaticais cumpridas pelas palavras nele contido, ao se relacionarem com outra palavra em um determinado sentido de associação são as seguintes:

- Nome ou substantivo (indicado pelo conjunto de caracteres [n],
- Adjetivo (indicado pelo conjunto de caracteres [a],
- Verbo (indicado pelo conjunto de caracteres [v] e
- Outros indicados pelo conjunto de caracteres [o]. Esta classe representa qualquer outra classe gramatical que não nome, verbo e adjetivo.

O *thesaurus* especifica também muitos casos em que palavras estão associadas a outras cumprindo mais de uma classe. Neste caso, o repositório apresenta classes duplas. Estas classes são as seguintes:

- [o/v], classe dupla, onde a palavra tem o papel de classe gramatical de verbo e de outros,
- [a/o], classe dupla, onde a palavra tem o papel de classe gramatical de adjetivo e de outros,
- [n/v], classe dupla, onde a palavra tem o papel de classe gramatical de verbo e de substantivo,
- [a/n], classe dupla, onde a palavra tem o papel de classe gramatical de adjetivo e de substantivo,
- [n/o], classe dupla, onde a palavra tem o papel de classe gramatical de nome e de outros,
- [v/a], classe dupla, onde a palavra tem o papel de classe gramatical de adjetivo e de verbo e,
- [a/v], classe dupla, onde a palavra tem o papel de classe gramatical de verbo e de adjetivo.

Pode-se notar que, [v/a] e [a/v] indicam essencialmente a mesma dupla classe, contudo como o *thesaurus* utilizou símbolos diferentes na sua confecção, esta diferente simbologia tem que ser levada em conta em termos de processamento de dados caso se deseje filtrar as palavras pelas suas classes gramaticais cumpridas nessas associações com outras palavras, o que foi especialmente levado em consideração nos desenvolvimentos na linguagem Java.

O motivo da escolha desse *thesaurus* reside inicialmente no fato de que foi este o único repositório aberto deste tipo (e que, portanto podia ser utilizado amplamente em rotinas computacionais e algoritmos) que se teve notícia durante boa parte do desenvolvimento deste trabalho. Somente no final dos desenvolvimentos efetuados é que se obteve notícia de que o *thesaurus* Wordnet também possuía uma versão aberta, sobre o qual os demais trabalhos efetuados na área de traços de personalidade e demais tinham feito uso.

Apesar desta informação, não se conseguiu encontrar tal versão, além do que também não se conseguiu obter informações adicionais sobre tal versão aberta (tais como legalidade de seu acesso, a cobertura dessa versão em termos de quantidade de palavras disponíveis etc.).

Finalmente, o banco adquirido e utilizado neste trabalho mostrou-se altamente compatível com as expectativas a cerca de um *thesaurus* aberto, contando inclusive com o grande atrativo de ser disponibilizado em diversos formatos; inclusive em planilha Excel, o que se mostrou altamente atrativo, pois a facilidade de acesso aos relacionamentos entre palavras e até mesmo de fácil modificação do *thesaurus* (com a inclusão e exclusão de palavras associadas a outras constantes no banco de dados) foram características muito importantes para o desenvolvimento de rotinas computacionais utilizando esta fonte de dados.

Mesmo levando-se em conta que algumas palavras procuradas, alguns traços de personalidade propostos por Catell, 1943b, não foram encontradas no *thesaurus* utilizado, esse fato não foi impeditivo dos desenvolvimentos apresentados a seguir, pois palavras próximas em significado foram encontradas como substitutas e, alternativamente, como mencionado acima) se poderia facilmente alterar o repositório original pela inclusão de palavras e seus relacionamentos com as demais), procedimento que seria amplamente facilitado pelo formato em planilha Excel.

5.2 LINGUAGENS DE PROGRAMAÇÃO UTILIZADAS

Todos os desenvolvimentos necessários, tanto do sistema importador de dados e gerador de relatórios como das demais rotinas e algoritmos desenvolvidos neste documento, foram feitos nas linguagens de programação PHP versão 5.3.9, compilada para 64 bits, disponível em <http://www.anindya.com/> e na linguagem Java.

Os desenvolvimentos em Java foram feitos nos ambientes de desenvolvimento Netbeans IDE versão 7.1.2 (disponível em <https://netbeans.org/>) e Eclipse Kepler versão 4.3 (disponível em <http://www.eclipse.org/downloads/index-developer.php>).

5.3 BANCOS DE DADOS UTILIZADOS

Para subsidiar os sistemas criados foram utilizados dois bancos de dados, a saber:

- Postgres SQL versão 8.4: utilizado para subsidiar o sistema desenvolvido em php.
- My SQL *Workbench*, 64 bits, versão 5.2. CE: subsidiando o sistema em Java. Sete bancos gratuitos encontram-se disponível para *download* em <http://www.mysql.com/>.

5.4 OUTROS SOFTWARES UTILIZADOS

Além das linguagens e bancos de dados mencionados, utilizou-se também outros *softwares*; seja para a instalação do sistema PHP, seja para rodar algoritmos ou organizar as informações dos relatórios gerados. Estes *softwares* são:

- Apache *Http Server Monitor* versão 2.2.20 compilado em 64 bits: este servidor Http foi necessário para a implementação de uma interface gráfica do sistema em php, inicialmente desenvolvido. Este programa pode ser baixado em. <http://www.anindya.com/>.

- SQL Manager Lite for PostGres SQL da EMS versão 5.3.0.1: importante na instalação do sistema em php. Disponível em <http://www.sqlmanager.net/>

- Notepad ++ versão 6.1.1: editor de texto utilizado para analisar e organizar grandes relatórios gerados em extensão.txt.

- Microsoft Excel dos Offices 2010 Professional e 2007 Student: utilizados para visualização, organização e formatação de alguns relatórios gerados.

- Microsoft Word dos Offices 2010 Professional e 2007 Student: utilizados para a digitação deste documento.

- MATLAB R2010b 64 bits versão 7.11.0.584 (de 16 de agosto de 2010) e *Toolbox Neural Networks* versão 7.0 (da mesma data de lançamento): utilizados para as implementações de redes neurais de Kohonen. Alguns suportes de implementação foram obtidos nos sites <http://www.opencadd.com.br/site/content/home/index.php>, e <http://www.mathworks.com/matlabcentral/>.

- *Software R* versão 2.12.2 com o pacote Igraph para plotagem do gráfico do dos componentes conexos grafo subjacente às relações entre palavras e dos clusters identificados.

5.5 SISTEMA DESENVOLVIDO EM LINGUAGEM PHP

Com o intuito de se aplicar algoritmos de clusterização nos grafos associados às relações entre um determinado conjunto de palavras selecionadas a partir do TTT adquirido, fazia-se necessário desenvolver diversos relatórios que trouxessem informações úteis a respeito destes. Havia a necessidade de se combinar as diferentes informações presentes em diferentes relatórios (informando cada um, aspectos diferentes do grafo subjacente às ligações entre palavras deste banco de palavras) e de gerar relatórios que por sua vez seriam input de outros algoritmos, como é o caso do relatório de vetores, a ser apresentado mais adiante, que seria input da rede neural de Kohonen, para clusterização de palavras no *thesaurus*.

Neste sentido houve a necessidade de se utilizar diversas rotinas computacionais que gerassem tais relatórios.

Adicionalmente, desejava-se que estas rotinas pudessem ser utilizadas por usuários leigos em programação e desenvolvimento, o que gerava a necessidade de agrupá-las em interfaces gráficas de fácil uso.

Estas necessidades, em conjunto, levaram ao desenvolvimento de um sistema desenvolvido em linguagem Php. As demais rotinas foram implementadas em linguagem Java (além de outras rotinas em MATLAB).

Como se verá mais adiante, muitos dos algoritmos implementados em Java estão aptos a resolver problemas de grafos em geral, e não somente a problemas relacionados ao *thesaurus* citado. Isto porque, tudo o que se necessita para utilizá-los é uma lista de adjacência que respeite os formatos utilizados em arquivo de texto em extensão.txt. Uma vez preparada esta lista com o formato requisitado (a ser explicado mais adiante) é possível se aplicar os algoritmos aqui utilizados para qualquer grafo, desde que este seja um 1-grafo.

Mesmo havendo alguns pacotes computacionais com algumas soluções de grafos disponíveis (sobretudo para o *software* MATLAB, como o MATGRAPH) estes pacotes ou não geram todos os relatórios de que se necessita ou não dispõem de todos os algoritmos de que se necessitava, sobretudo com a flexibilidade necessária.

Adicionalmente, se temia que estes *softwares* não conseguissem lidar com grandes quantidades de dados (no que tange mais especificamente a aplicações com o TTT, a grandes quantidades de palavras utilizadas) que aqui se almeja.

Por todas estas considerações, houve a necessidade de se desenvolver um sistema, incorporando conjuntos de rotinas e capaz de gerar alguns relatórios úteis para os problemas aqui abordados, sobre interfaces gráficas amigáveis, que pudessem gerar os relatórios necessários importantes como input para outros algoritmos, de forma acessível a usuários não desenvolvedores.

Por sistema, se estará denotando aqui, um conjunto de rotinas computacionais desenvolvidas em linguagem PHP, reunidas sob uma mesma GUI (*graphic user interface*) que executam rotinas relacionadas a grafos, gerando relatórios ao seu final.

No que diz respeito aos algoritmos de grafos na linguagem Java aqui utilizado, na grande maioria dos casos, foram baixados de *sites* específicos da *internet*, testados exaustivamente para diversos conjuntos de dados e posteriormente modificados e aperfeiçoados para atenderem os seguintes quesitos:

- Gerarem relatórios automaticamente,
- Lidarem com grandes quantidades de dados.

Este capítulo apresentará o sistema acima citado e demais implementações algorítmicas desenvolvidas para atender as necessidades de geração de relatórios em formatos específicos desta tese, e tratará de todas as seus utilitários e características.

5.5.1 Sistema PHP

Este sistema foi elaborado na fase inicial de desenvolvimento deste trabalho, quando o *thesaurus* ainda não havia sido adquirido (trabalhava-se somente com uma versão de amostra gratuita disponível para *download* da página do fabricante do mesmo) e não se tinha ideia da quantidade de palavras disponíveis na versão completa do mesmo, na língua inglesa.

Escolheu-se a linguagem PHP para sua confecção, pois, face às limitações de *hardware* presentes no início dos desenvolvimentos deste documento (inicialmente os primeiros desenvolvimentos foram feitos com um *notebook* com CPU Intel Core 2 Duo e sistema operacional Windows Vista 32 bits), aliado ao julgamento equivocado de que o número de palavras na versão completa do *thesaurus* não seria muito maior do que na versão gratuita) tal linguagem facilitava o acesso ao sistema por meio da própria interface de programas (*browsers*) navegadores da *internet* (no caso, do Google Chrome). O sistema, nesta ocasião se encontrava inicialmente hospedado em um servidor gratuito.

Adicionalmente, os desenvolvimentos nesta linguagem eram muito mais fáceis e rápidos de serem efetuados do que na linguagem Java.

Este sistema é formado, como já esclarecido, em linguagem PHP, tendo também uma interface gráfica (GUI) em HTML. Ele é composto basicamente dos seguintes componentes:

- O sistema foi desenvolvido em linguagem PHP,
- Necessita de um servidor HTTP, no caso utilizou-se o servidor Html Apache versão 2.2.20,
- Acessa dados importados no banco de dados POSTGRES SQL.

Posteriormente, com a consideração de todos os adjetivos do *thesaurus* (o que será visto no próximo capítulo) e da descoberta de novos e melhores servidores e serviços de *cloud computing*, este sistema deixou de poder gerar os relatórios necessários e se desenvolveu rotinas em linguagem Java para melhor atender usos intensivos (com grandes quantidades de palavras) das mesmas.

Este sistema é capaz de, para pequenas quantidades de dados, importar partes do *thesaurus* (respeitando-se os mesmos formatos originais neste banco de palavras) e gerar relatórios os seguintes relatórios:

- Matriz de Adjacência
- Lista de Adjacência
- Relação de Palavras e Rótulos Numéricos
- Relatório de Representação das palavras como Vetores em Alta Dimensão

Estes relatórios serão posteriormente explicados.

Quando finalmente se adquiriu o *thesaurus* completo tal sistema deixou de atender as necessidades do trabalho (pelas próprias limitações da linguagem, oficialmente fornecida somente na versão 32 bits pelo seu fabricante). Mesmo assim, por se depender deste sistema para gerar os relatórios de representação das palavras do TTT como vetores, houve a necessidade de se reconfigurar o mesmo para abrigar as versões (não oficiais) das linguagens PHP e do servidor Apache em 64 bits.

Posteriormente, com a disponibilidade de melhores *hardwares*, servidores e serviços de *cloud computing* para a continuidade dos desenvolvimentos deste trabalho, este sistema deixou de ser utilizado para gerar os relatórios necessários e foi substituído pelas funcionalidades em linguagem Java (com exceção da funcionalidade relacionada à codificação de palavras como vetores, que continuou a ser feita exclusivamente no sistema modificado para abrigar as versões não oficiais de PHP e Apache em 64 bits).

Assim, neste sistema, a parte da planilha em Excel contendo as palavras desejadas (se apagava da planilha todas as demais palavras, por limitações do sistema em lidar com grandes quantidades das mesmas, o mesmo permanecendo verdade na versão adaptada em 64 bits) era lida e importada para o banco de dados POSTGRES (estruturado).

Esta importação gerava uma lista, no sistema PHP, que deveria ser nomeada pelo usuário. Desta lista se poderiam gerar os relatórios mencionados, ou seja, e a partir das importações das palavras o sistema gerava uma lista, contendo as mesmas, e destas listas o sistema gerava relatórios.

Este sistema é capaz de gerar os seguintes relatórios:

- **Listas de palavras e seus rótulos numéricos associados:** esse relatório atribui, para cada palavra, um rótulo numérico único, que pode representá-la, nos demais relatórios produzidos por este sistema.

-**Listas de adjacência:** esse relatório é um importante método de representação das estruturas de um grafo. Ele relaciona, para cada palavra do banco de palavras (que é representada como um nó de um grafo, como se verá mais adiante) todas as palavras

diretamente relacionadas a ela. Este relatório pode representar estas associações usando as palavras em si ou utilizando os rótulos numéricos que as representam.

- **Matrizes de Adjacência:** este é outro relatório importante a cerca de um grafo. Neste relatório, representa-se por meio de uma matriz composta de zeros e uns (ou como uma matriz esparsa) as presenças ou ausências de ligações entre cada par de nós em um grafo.

No caso deste sistema PHP/POSTGRES, a matriz criada é composta por zeros (para indicar ausência de ligações) e uns (para representar a presença de ligações) entre pares de palavras. Contudo, por características e limitações no gerenciamento da memória do computador, este sistema só é capaz de gerar matrizes de pequenas dimensões (no máximo com algumas centenas de linhas e colunas).

- **Relatório de codificação das palavras como vetores de alta dimensão:** este relatório codifica as palavras do banco de dados como vetores de alta dimensão de forma biunívoca (ou seja, a cada palavra há a sua representação como um vetor único), como algoritmo apresentado anteriormente. As rotinas de preparo deste relatório foram desenvolvidas para atenderem as necessidades de codificação das palavras inicialmente concebidas para treinamento das redes SOM.

5.5.2 Interfaces gráficas do sistema em PHP

A seguir se apresentará suas funcionalidades por meio das telas do sistema PHP.



Figura 24: Tela de Entrada do Sistema PHP

Na tela de entrada do sistema, tem-se, no menu à esquerda as opções Importar Dados (planilhas Excel no formato do TTT), Listas, Gerador de Matriz, Relatórios e *Download*, a serem explicadas nesta sessão.

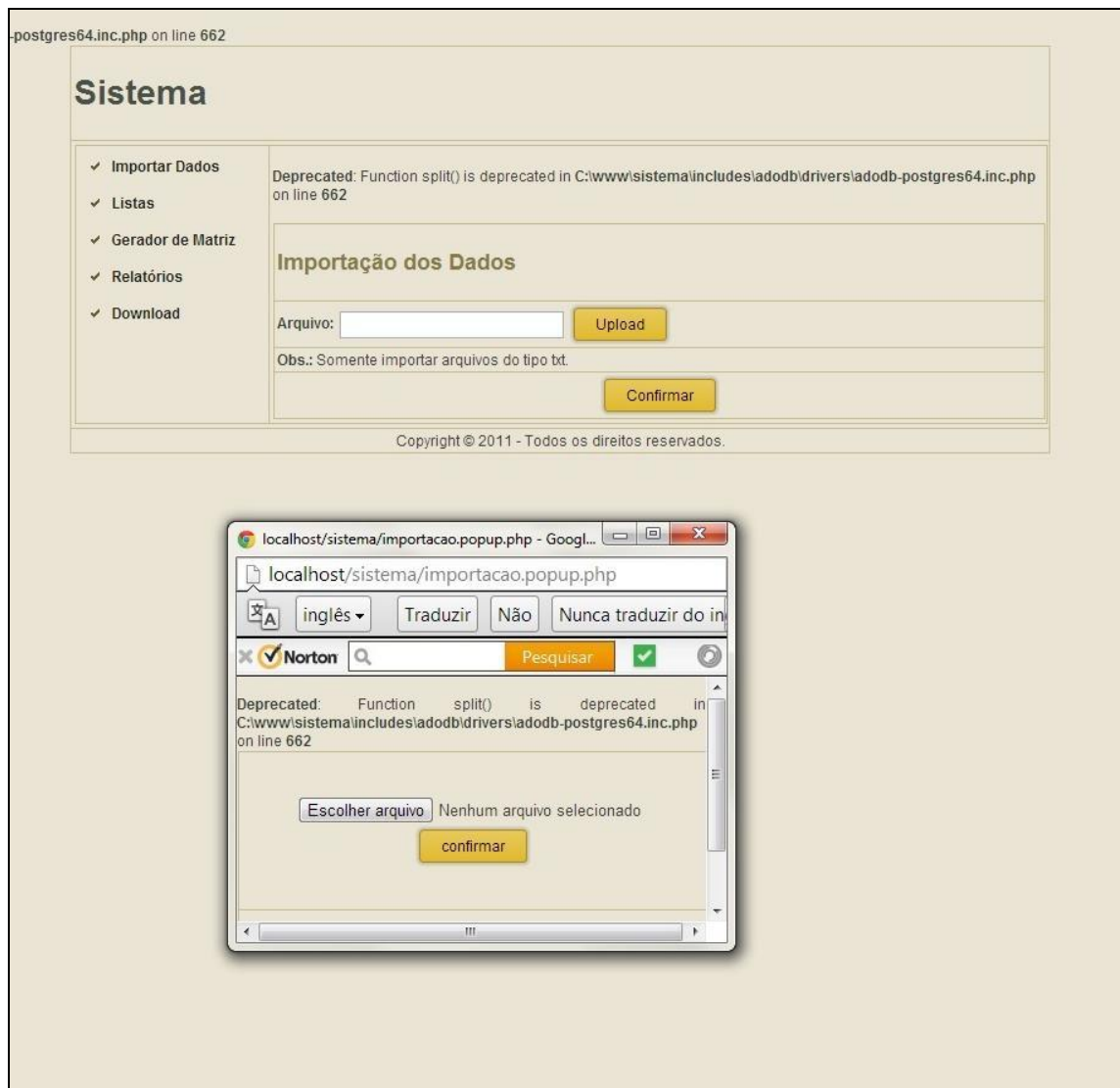


Figura 25: Tela de Importação de Palavras do Sistema em PHP

Na tela de importação de dados, pode-se importar o *thesaurus* na versão em planilha de Excel (mais especificamente partes dele), clicando-se no botão *Upload*. Ao fazê-lo, abre-se uma janela *pop up* em que se deve pressionar o botão escolher arquivo. Uma vez pressionado este botão, o caminho da planilha Excel contendo o *thesaurus* deve ser especificado.

postgres64.inc.php on line 662

Sistema

- ✓ Importar
- ✓ Listas
- ✓ Gerador
- ✓ Relatórios
- ✓ Download

Deprecated: Function split() is deprecated in C:\www\sistema\includes\adodb\drivers\adodb-postgres64.inc.php on line 662

Listas Geradas

Importação	Lista	Ordem de Relacionamento	Qty Keywords Primárias	Qty Keywords Secundárias	Qty Palavras	Incluir
Importação - 02/07/2013 11:07:01	SIN_ENG_227KWDS_FINAL_02-07-20131	MAX	227	0	227	Excluir
Importação - 02/07/2013 09:07:43	SIN_ENG_227KWDS_FINAL_02-07-2013	MAX	227	0	227	Excluir
Importação - 29/06/2013 12:06:34	228 entre elas 29-06-2013	MAX	228	0	229	Excluir
Importação - 08/06/2013 10:06:12	228_1_ORDEM_FINAL	MAX	270	1598	2669	Excluir
Importação - 01/06/2013 06:06:42	228_SO_ELASFINALFINAL	MAX	228	0	229	Excluir
Importação - 28/05/2013 11:05:07	228_SO_ELASFINAL111	MAX	228	0	229	Excluir
Importação - 28/05/2013 11:05:26	228_SO_ELASFINAL11	MAX	228	0	229	Excluir
Importação - 28/05/2013 11:05:29	228_SO_ELASFINAL1	MAX	227	0	228	Excluir
Importação - 28/05/2013 11:05:29	228_SO_ELASFINAL	MAX	227	0	228	Excluir
Importação - 28/05/2013 04:05:21	228_SO_ELAS	MAX	227	0	234	Excluir

1 - 2

Copyright © 2011 - Todos os direitos reservados.

Figura 26: Tela de Listas no Sistema em PHP

Na tela de listas é possível obter informações sobre as listas já criadas com base em importações passadas efetuadas pelo sistema. Assim, na coluna Importação tem-se a data e hora em que ocorreu a importação da lista considerada. Na coluna Lista tem-se o nome dado às listas criadas.

Na coluna Ordem de Relacionamento, tem-se o número máximo de ordem de relacionamento (isto é, à distância) que as palavras importadas devem ter, em relação às suas *keywords*, para serem consideradas para confecção de um determinado relatório. As demais colunas referem-se à quantidade de *keywords* presentes em cada lista gerada e o total de palavras em cada lista.

Caso se deseje criar uma nova lista, a partir das importações passadas efetuadas pelo sistema, se deverá clicar no botão Incluir, o que remeterá o usuário a uma nova tela, apresentada a seguir.

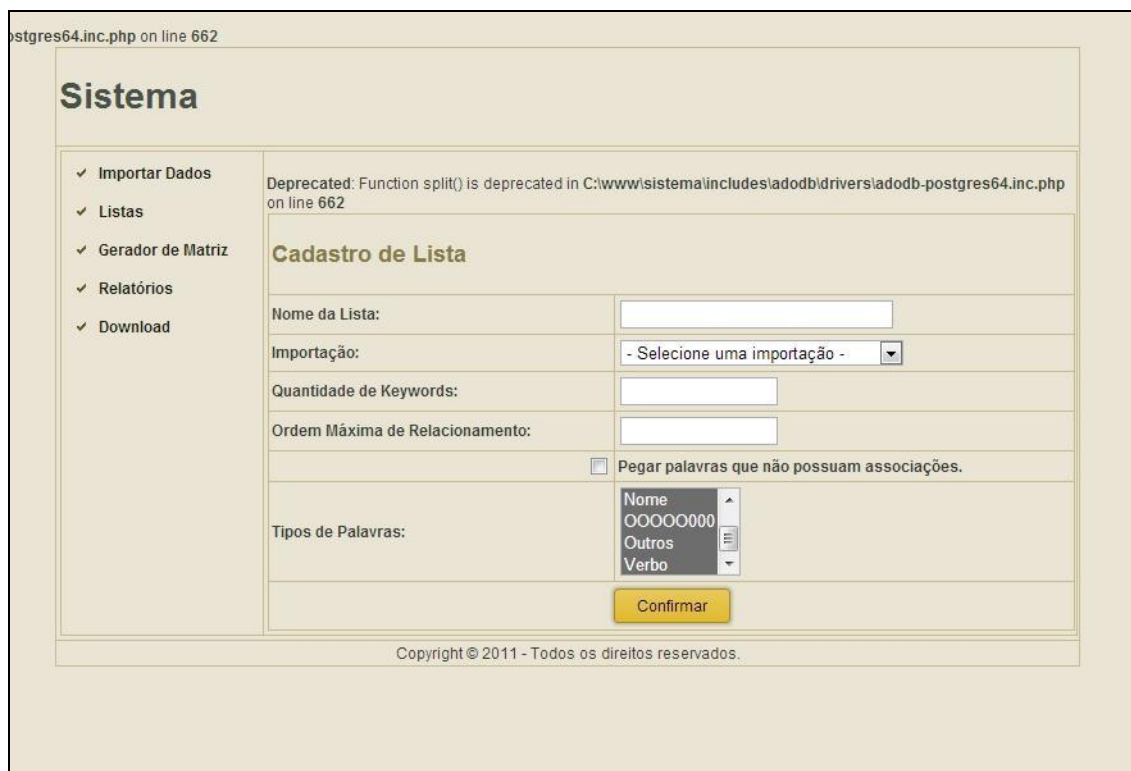


Figura 27: Tela de Criação de Listas no Sistema em PHP

Na tela de cadastro de uma nova lista deve-se especificar um nome para a lista nova a ser criada. Na aba de importação deve-se selecionar qual importação (identificada por data e hora) será utilizada para compor a nova lista.

O campo Quantidade de *Keywords*, se refere ao número de *keywords* (e as respectivas palavras a elas associadas, ao longo das outras colunas da planilha), lidas de cima para baixo na planilha, a serem consideradas, em uma dada importação, para a confecção de uma lista.

O campo Ordem Máxima de Relacionamento (que na realidade são as distâncias euclidianas entre os vetores que representam cada uma das palavras consideradas) se refere à ordem máxima de relacionamento que as palavras (que não sejam *keywords* primárias) devem ter em relação a estas para serem consideradas na confecção de um dado relatório.

Se o campo Ordem Máxima de Relacionamento for deixado em branco todas as palavras, ligadas direta ou indiretamente (isto é, em todas as ordens de relacionamento) às *keywords* selecionadas (primárias) presentes na importação, serão consideradas.

Finalmente, o campo Tipos de Palavras refere-se às classes gramaticais (classes essas cumpridas pelas palavras, ao assumirem um dado sentido de associação com uma *keyword*) que podem ser selecionadas para a confecção de um dado relatório. Assim podem-se filtrar palavras, para serem incluídas ou não nos relatórios a serem gerados, com base nessas classes gramaticais (por exemplo, em uma aplicação psicoléxica pode-se ter interesse em se

considerar somente associações entre palavras em que estas representem somente a classe de adjetivo).

Neste campo, têm-se as classes Nome (classe n no dicionário), Adjetivo (classe a no banco de palavras), Verbos (classe v) e Outros (classe o). Tem-se ainda uma opção identificada pela sequência de caracteres “OOOOO000” que representa todas as classes duplas presentes no banco (isto é, as classes o/v, a/o, n/v, a/n, n/o, v/a e a/v). Assim, ao se selecionar esta opção, todas as classes duplas serão consideradas conjuntamente.



Figura 28: Tela de Geração de Matriz de Adjacência do Sistema em PHP

Nesta tela deve-se selecionar a lista a partir da qual se deseja gerar uma matriz de adjacência para o grafo associado às relações entre as palavras consideradas na lista.

Uma vez selecionada a lista deve-se clicar no botão Gerar Estimativa. Este botão irá acionar um relatório que irá indicar quantos arquivos.txt em separado serão gerados (a geração desta matriz, neste sistema em linguagem PHP é feita por blocos; isto é, cada arquivo gerado em separado corresponde a um conjunto de linhas da matriz de adjacência. Depois de gerados todos estes arquivos o sistema os concatena em um único arquivo.txt – representando toda a matriz de adjacência – e disponibiliza este arquivo completo para o usuário).

Tal procedimento de geração da matriz por blocos de linhas foi necessário devido ao esforço computacional para gerar este relatório na linguagem PHP (especialmente na versão original de 32 bits) o que, na versão original do sistema (em 32 bits) enfrentava problemas de travamento de processamento por falta de memória RAM já com alguns milhares de palavras consideradas.

A matriz de adjacência gerada nesta tela é composta por zeros e uns. O número zero é utilizado para indicar ausência de adjacência entre duas palavras e o número um é utilizado para indicar presença de adjacência entre duas palavras.



Figura 29: Tela de Geração dos Demais Relatórios do Sistema em PHP

Ao se selecionar a opção Relatórios no menu do sistema o mesmo apresenta a tela acima exposta. Nesta tela, o usuário definirá a lista a partir da qual irá gerar os demais relatórios e o tipo de relatório que deseja. Estes relatórios são: Lista de Adjacência, Relação *Label*/Palavra, e Relatório de Vetor (em extensão.txt ou.csv), já explicados anteriormente.

Adicionalmente, estes relatórios (com exceção dos relatórios de vetores) podem apresentar as relações entre as palavras representadas por elas mesmas (isto é, pela sequência de caracteres que representa cada palavra) ou pelos rótulos numéricos (*labels*) que as representam. A figura a seguir apresenta o menu Tipo de Relatório expandido.



Figura 30: Tela de *Download* dos Relatórios Gerados pelo Sistema em PHP

A figura acima apresenta a tela de Relatórios do sistema com a aba Tipo Relatório expandida.

5.5.3 Exemplos de Relatórios Gerados Pelo Sistema em PHP

No Apêndice desta tese, encontram-se alguns exemplos de relatórios gerados por este sistema.

5.5.4 Melhoramentos possíveis de serem feitos no sistema em PHP

Ao longo de sua confecção e utilização detectaram-se alguns aspectos que poderiam ser aperfeiçoados no sistema em PHP desenvolvido. Em

- filtro por ordem de relacionamento das palavras em relação às *keywords* com as quais se relacionam diretamente: este filtro se baseia na distância entre estas palavras, medida somente segundo a distância entre os vetores de alta dimensão que as representam. Este procedimento só representa de forma totalmente precisa todas as distâncias entre todas as palavras consideradas quando o grafo representando as relações entre elas for do tipo árvore ou floresta (caso contrário se estará errando a distância entre palavras pertencentes a ciclos, neste grafo).

Quando não se pode garantir a priori que o grafo representante das ligações entre palavras selecionadas for árvore ou floresta, não se deve utilizar este filtro.

Além da verificação da inadequabilidade deste sistema PHP para lidar com grandes quantidades de palavras (principalmente por conta das limitações da própria linguagem), verificou-se, adicionalmente, alguns aspectos que deveriam ser aperfeiçoados no mesmo (estes aspectos também se constituíram em algumas das motivações para o desenvolvimento do sistema na linguagem Java):

- Ausência de filtros individualizados para as palavras com classes duplas: para o caso de palavras que se associam a outras cumprindo simultaneamente mais de uma classe gramatical, não se desenvolveu, neste sistema PHP, filtros individuais de seleção para cada uma dessas classes dupla. Na atual configuração todas as classes duplas foram reunidas em um único filtro, intitulado “OOOOO000”, como visto anteriormente. Esta deficiência foi sanada no desenvolvimento do sistema em Java.

- Processo de geração de matriz de adjacência pouco eficiente: em função tanto da linguagem escolhida como do algoritmo implementado para tal função, o processo de geração de matrizes de adjacência neste sistema é pouco eficiente (por esta razão, optou-se pela geração desta matriz por blocos, onde cada bloco corresponde a um determinado

conjunto de linhas da mesma. Ao final da geração de todos os blocos se constrói a matriz toda pela concatenação desses blocos).

- Melhoramentos a serem feitos na funcionalidade relacionada à construção de listas de adjacência: observou-se ainda que os relatórios de listas de adjacência geradas por este sistema apresentavam duas características que poderiam ser aperfeiçoadas (e o foram no sistema seguinte). Estas características são:

a) As listas de adjacência só incluem palavras que possuem adjacência com pelo menos mais outra palavra. Palavras isoladas não são incluídas neste relatório.

b) Para o caso de palavras pertencentes a ciclos (no grafo correspondente à rede de ligações entre as palavras selecionadas), há repetições de palavras no relatório, no lado direito da linha separatriz da lista de adjacência (isto é, no lado em que há a listagem de todas as palavras diretamente relacionadas a uma dada palavra especificada no lado esquerdo, antes desta linha). Estas repetições devem ser eliminadas manualmente. Este problema também foi elei minado no sistema seguinte desenvolvido.

Deve-se esclarecer aqui que estas características presentes nos relatórios de lista de adjacência não influenciam a confecção dos demais relatórios preparados por este sistema.

- O sistema PHP só produz relatórios para palavras presentes no *thesaurus* utilizado: diferentemente das rotinas implementadas em Java, apresentadas logo a seguir, o sistema aqui apresentado só é capaz de gerar relatórios para palavras presentes no TTT adquirido. As rotinas em Java, por sua vez, eliminaram estas limitações, apresentando uma série de interfaces gráficas reunindo rotinas relacionadas a grafos (e a determinação de *clusters* em grafos) que geram relatórios a partir de outros relatórios de entrada (*input*).

- Os vetores gerados pelo sistema PHP possuem como coordenadas não nulas para as componentes adicionais, introduzidas no final do processo de codificação - para dar conta de atribuição de distâncias muito grandes entre vetores pertencentes a diferentes componentes conexos do grafo – somente componentes assumindo valores fixos iguais a 500.

Isto implica que, depois de gerados em relatórios.txt esses vetores terão que ter os valores dessas componentes não nulas alteradas manualmente de forma que vetores pertencentes a componentes conexas diferentes tenham valores diferentes para as mesmas e ao mesmo tempo em que vetores pertencentes ao mesmo componente conexo tenham valores idênticos para estas coordenadas.

Como se verá todos estes relatórios depende (direta ou indiretamente) de um relatório-base, que é a lista de adjacência. Esta lista de adjacência pode ser tanto produzida a partir de

palavras presentes no *thesaurus* em questão, como fornecidas independentemente do mesmo (desde que o usuário obedeça aos formatos de arquivo requeridos).

Com isto, todas as funcionalidades estão disponíveis para qualquer problema envolvendo 1-grafos e não mais restritas a aplicações com *thesauri*.

5.5.5 Rotinas em Java

Estas rotinas também foram desenvolvidas com os objetivos de gerar relatórios para as tarefas desenvolvidas neste trabalho, porém com performances computacionais muito superiores às do sistema em PHP, podendo ser aplicadas em grande parte ou todas as palavras do TTT.

Elas se diferenciam das do sistema anterior fundamentalmente na escolha da linguagem e nas otimizações de código para que as gerações desses relatórios pudessem se dar com qualquer quantidade de palavras, incluindo o TTT inteiro, caso necessário; assim como fazer o uso mais otimizado possível da memória e processamento do computador.

Adicionalmente contam com as seguintes características:

- Recursos de portabilidade da linguagem Java (podendo rodar em qualquer arquitetura de *hardware* e qualquer sistema operacional sem necessidade de reprogramação ou novos desenvolvimentos por parte do usuário. Esta portabilidade dessa linguagem é uma característica importante para os usuários potenciais das ferramentas implementadas, especialmente em casos onde um *hardware* inicialmente utilizado não atenda as necessidades de processamento e memória das rotinas) e haja necessidade de migração para outro, contando com outro sistema operacional,

- Estas rotinas também produzem um número muito maior de relatórios ligados a rotinas de grafos do que o sistema anterior (só não produzindo o relatório de vetores).

- Cada rotina implementada em Java depende de um relatório específico de *input* para gerar outro tipo de relatório. Assim, dependendo do tipo de relatório que o usuário deseje desenvolver ele terá que gerar um ou mais relatórios anteriores, por rotinas diferentes para finalmente gerar o seu relatório final de interesse.

- Como se verá adiante, todos os demais relatórios produzidos por estas rotinas, dependem fundamentalmente (direta ou indiretamente) de um relatório base: a lista de adjacência do grafo. Esta lista pode ser feita manualmente desde que respeite os formatos requisitados para servirem de *input* à rotina de geração desta lista para gerar outros tipos de relatórios.

Desta forma, todos os recursos e algoritmos aqui implementados em Java (relacionados a grafos e também o algoritmo de clusterização em grafos conhecido como MCL) ficam aplicáveis a qualquer problema de grafo (inclusive não relacionados a *thesauri*), desde que os grafos considerados sejam 1-grafos.

Alternativamente, em problemas específicos de grafos (que demandem soluções como as desenvolvidas e/ou implementadas neste trabalho) relacionados às palavras contidas no TTT utilizado, pode-se gerar tal relatório base (lista de adjacência) a partir da importação da planilha contendo o banco mencionado (todo, ou parte dele).

Nestes casos, a planilha é importada para um banco de dados MySQL (64-bits versão 5.2 CE) e deste banco de dados se podem gerar (de forma customizada) tal relatório.

- As rotinas foram implementadas em parte na interface IDE Netbeans e em parte na interface Eclipse.

5.5.5.1 Rotina de Importação de Dados do TTT em Planilha Excel

A primeira rotina desenvolvida é a que se encarrega da importação de palavras da planilha. xls (toda ou partes dela) contendo o *thesaurus* para o banco de dados MySQL.

5.5.5.2 Rotina de geração de listas de adjacência a partir de dados importados

Nesta rotina, têm-se as funcionalidades referentes à geração de listas de adjacência a partir de importações feitas anteriormente pelo usuário. Ela demanda do usuário que o mesmo especifique o nome do arquivo que a lista de adjacência receberá (a lista de adjacência tem o formato de arquivo de texto, separado por tabulações).

Ela também permite filtragens de palavras, a serem incluídas na lista de adjacência, baseadas nas classes gramaticais que estas desempenham ao se relacionarem com outras palavras no *thesaurus* (segundo um determinado sentido de associação).

Esta rotina também engloba funcionalidades para que o usuário construa listas de adjacência contendo somente que palavras especificadas por ele (caso existam), em uma determinada importação.

Assim a rotina checa se tais palavras existem ou não em uma dada importação de palavras, formando a lista com as palavras de interesse existentes nesta mesma importação.

Esta rotina também gera um arquivo de *log*, onde constará qual (is) palavra(s) não foi (foram) encontrada(s) na seleção feita pelo usuário.

Esta funcionalidade fornece ao potencial usuário grande flexibilidade de trabalho e produtividade nas suas atividades, pois:

- Automaticamente faz a checagem, para cada palavra digitada, se a mesma está presente na importação de dados.

- Faz a mesma Checagem para as classes gramaticais selecionadas.

Tais expedientes poupam o usuário de ter que fazer procuras no arquivo do TTT para constatar se determinadas palavras (e classes a elas associadas) de interesse estão presentes no banco ou não, manualmente ou por meio de ferramentas de busca pouco desenvolvidas, como por exemplo, as do *software* Excel.

5.5.5.2.1 Programação Orientada a Objeto no Banco de Dados MySQL

Visando-se assegurar de que não haveria erros na importação das palavras, nas identificações de suas relações de adjacência e nas classes gramaticais por elas cumpridas, se concebeu uma forma de representação destas palavras e dos seus diferentes atributos, no TTT, por meio de programação orientada a objeto.

Neste sentido, representou-se cada palavra e todas as demais informações a cerca da mesma, no TTT (suas relações de adjacência, se a mesma é *keyword* ou não, a classe gramatical por ela exercida em cada uma de suas adjacências com outras palavras, o sentido que a associa com outras palavras) por meio de um conjunto de classes (no sentido de orientação a objeto). Tal procedimento também visou obter maior eficiência computacional na elaboração deste relatório.

Esta representação só foi possível por meio de programação orientada a objeto (o que se efetuou dentro do banco de dados estrutural *My SQL*) e está ilustrada na figura abaixo:

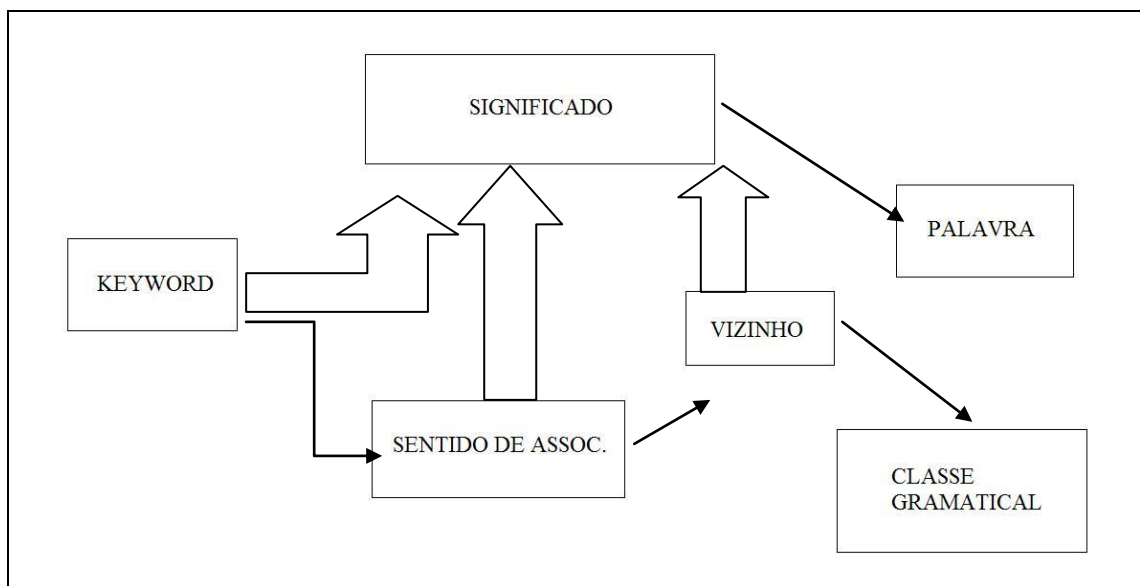


Figura 31: Diagrama Conceitual de Classes Utilizadas na Programação a Objeto no Banco de Dados *My SQL*

Na figura acima, verifica-se que os desenvolvimentos orientados a objeto feitos no banco de dados utilizou as seguintes classes concretas: *keyword*, sentido de associação, vizinho, classe gramatical e palavra. Utilizou também a classe abstrata significado.

As setas cheias, na figura, representam as relações entre classes concretas e a classe abstrata significado.

No diagrama conceitual utilizado, uma palavra no TTT é referenciada por uma classe abstrata denominada significado. Esta última, por ser uma classe abstrata, assume os possíveis “estados” *keyword*, sentido de associação e vizinho.

A classe *keyword* está ligada à classe sentido de associação. Esta última representa um (ou mais) vizinho(s). A classe vizinho por sua vez se relaciona a uma classe gramatical.

Desta forma, de forma não técnica, pode-se interpretar que uma palavra pode assumir diferentes significados: ela pode ser *keyword* no TTT, pode ser vizinha de uma *keyword*, e pode assumir um sentido que a associa com uma ou mais palavras (sentido de associação).

Caso seja vizinha, terá que assumir, para cada relação de vizinhança com outra palavra, o papel de uma classe gramatical (segundo as normas com as quais o *thesaurus* utilizado foi organizado).

O *output* desta rotina, como já mencionado é a lista de adjacência contendo as palavras (e classes suas gramaticais) selecionadas. A seguir se apresenta um exemplo de lista da adjacência gerada por esta rotina (formatada em termos de tabulação). As palavras contidas nesta lista correspondem às mesmas 222 utilizadas nas aplicações feitas mais adiante neste trabalho (destas 222 palavras, as que não possuíam associação com nenhuma outra desta seleção, não são incluídas na lista).

Tabela 21: Exemplo de lista de adjacência gerada por interface de geração de lista de adjacência em java com 222 palavras

<i>excitable</i>	<i>enthusiastic,high-strung,irritable</i>
<i>sentimental</i>	<i>affected,emotional,idealistic</i>
<i>kind-hearted</i>	<i>amiable,affectionate,responsive</i>
<i>amiable</i>	<i>social,charming,genial,kind,kind-hearted</i>
<i>charming</i>	<i>Amiable</i>

<i>kind</i>	<i>amiable,friendly</i>
<i>devout</i>	<i>pious,reverent,heartly,honest,religious</i>
<i>honest</i>	<i>devout, Frank</i>
<i>frank</i>	<i>honest,open-hearted</i>
<i>open-hearted</i>	<i>Frank</i>
<i>religious</i>	<i>devout,pious,reverent</i>
<i>pious</i>	<i>religious,devout,reverent</i>
<i>affected</i>	<i>conceited,sentimental</i>
<i>reverent</i>	<i>pious,religious,devout</i>
<i>optimistic</i>	<i>Cheerful</i>
<i>affable</i>	<i>sociable,social,grateful</i>
<i>grateful</i>	<i>Affable</i>
<i>political</i>	<i>Social</i>
<i>informative</i>	<i>hospitable,social,polished</i>
<i>hospitable</i>	<i>informative,polished,social</i>
<i>poised</i>	<i>Polished</i>
<i>eclectic</i>	<i>many-sided,multifarious,discriminating</i>
<i>many-sided</i>	<i>multifarious,multiple,eclectic</i>
<i>conceited</i>	<i>arrogant,boastful,affected,egotistical,imitative,self-satisfied</i>
<i>multifarious</i>	<i>many-sided,eclectic</i>
<i>multiple</i>	<i>many-sided</i>

<i>constructive</i>	<i>Practical</i>
<i>economical</i>	<i>frugal, provident, sparing, thrifty, practical</i>
<i>frugal</i>	<i>economical, sparing, provident, thrifty</i>
<i>provident</i>	<i>frugal, sparing, thrifty, economical</i>
<i>thrifty</i>	<i>frugal, provident, economical, sparing</i>
<i>sparing</i>	<i>thrifty, economical, frugal, provident</i>
<i>vivacious</i>	<i>enthusiastic, active</i>
<i>witty</i>	<i>Clever</i>
<i>arrogant</i>	<i>autocratic, boastful, egotistical, conceited</i>
<i>mindful</i>	<i>cautious, thoughtful</i>
<i>cautious</i>	<i>Mindful</i>
<i>academic</i>	<i>speculative, conjectural</i>
<i>conjectural</i>	<i>academic, speculative, doubtful</i>
<i>doubtful</i>	<i>conjectural, self-conscious</i>
<i>self-conscious</i>	<i>Doubtful</i>
<i>imaginary</i>	<i>ideal, subjective</i>
<i>subjective</i>	<i>Imaginary</i>
<i>anxious</i>	<i>panicky, worried, afraid, enthusiastic</i>
<i>worried</i>	<i>anxious, frightened</i>
<i>boastful</i>	<i>arrogant, conceited</i>
<i>frightened</i>	<i>Worried</i>

<i>panicky</i>	<i>anxious, afraid</i>
<i>afraid</i>	<i>anxious, panicky, timid</i>
<i>timid</i>	<i>afraid, coward</i>
<i>coward</i>	<i>Timid</i>
<i>high-strung</i>	<i>Excitable</i>
<i>autocratic</i>	<i>Arrogant</i>
<i>prominent</i>	<i>curious, leading</i>
<i>curious</i>	<i>interested, prominent, visible</i>
<i>visible</i>	<i>curious, perceivable, physical</i>
<i>physical</i>	<i>corporeal, corporal, visible</i>
<i>corporeal</i>	<i>animal, bodily, corporal, physical</i>
<i>animal</i>	<i>bodily, corporeal</i>
<i>bodily</i>	<i>animal, corporal, corporeal</i>
<i>corporal</i>	<i>physical, bodily, corporeal</i>
<i>perceivable</i>	<i>Visible</i>
<i>interested</i>	<i>Curious</i>
<i>egotistical</i>	<i>conceited, self-satisfied, selfish, arrogant, opinionated</i>
<i>leading</i>	<i>Prominent</i>
<i>argumentative</i>	<i>Pugnacious</i>
<i>pugnacious</i>	<i>Argumentative</i>
<i>assertive</i>	<i>Vocal</i>

<i>vocal</i>	<i>speaking,unwritten,assertive</i>
<i>speaking</i>	<i>Vocal</i>
<i>unwritten</i>	<i>Vocal</i>
<i>self-satisfied</i>	<i>conceited,self-confident,egotistical</i>
<i>complaining</i>	<i>Cynical</i>
<i>cynical</i>	<i>complaining,sarcastic</i>
<i>sarcastic</i>	<i>Cynical</i>
<i>conscientious</i>	<i>tactful,painstaking</i>
<i>painstaking</i>	<i>Conscientious</i>
<i>tactful</i>	<i>Conscientious</i>
<i>controlled</i>	<i>Patient</i>
<i>patient</i>	<i>controlled,persevering</i>
<i>self-confident</i>	<i>self-satisfied</i>
<i>persevering</i>	<i>Patient</i>
<i>decorative</i>	<i>Musical</i>
<i>musical</i>	<i>Decorative</i>
<i>opinionated</i>	<i>Egotistical</i>
<i>headstrong</i>	<i>Mulish</i>
<i>mulish</i>	<i>Headstrong</i>
<i>enthusiastic</i>	<i>anxious,hearty,excitable,impulsive,vivacious</i>
<i>selfish</i>	<i>Egotistical</i>

<i>lexical</i>	<i>Verbal</i>
<i>verbal</i>	<i>Lexical</i>
<i>imitative</i>	<i>Conceited</i>
<i>reliable</i>	<i>Stable</i>
<i>stable</i>	<i>Reliable</i>
<i>idealistic</i>	<i>sentimental,visionary</i>
<i>visionary</i>	<i>idealistic,ideal</i>
<i>ideal</i>	<i>abstract,conceptual,hypothetical,visionary,imaginary</i>
<i>abstract</i>	<i>conceptual,hypothetical,ideal</i>
<i>hypothetical</i>	<i>abstract,conceptual,speculative,ideal</i>
<i>conceptual</i>	<i>abstract,hypothetical,ideal</i>
<i>speculative</i>	<i>academic,conjectural,hypothetical,thoughtful</i>
<i>thoughtful</i>	<i>observing,speculative,mindful</i>
<i>impulsive</i>	<i>emotional,enthusiastic,instinctive</i>
<i>observing</i>	<i>alert,intelligent,thoughtful</i>
<i>intelligent</i>	<i>alert,clever,logical,observing,rational,shrewd,wise</i>
<i>shrewd</i>	<i>clever,hard,intelligent</i>
<i>clever</i>	<i>intelligent,shrewd,witty>alert,dexterous,skilful</i>
<i>skilful</i>	<i>clever,deft,prepared,skilled,dexterous</i>
<i>deft</i>	<i>dexterous,skilful,skilled</i>
<i>skilled</i>	<i>deft,proficient,skilful</i>

<i>proficient</i>	<i>professional,skilled</i>
<i>professional</i>	<i>Proficient</i>
<i>dexterous</i>	<i>clever,skilful,deft</i>
<i>instinctive</i>	<i>impulsive,intuitive,mechanical</i>
<i>prepared</i>	<i>Skilful</i>
<i>alert</i>	<i>active,clever,fresh,observing,intelligent</i>
<i>active</i>	<i>alert,fresh,vivacious,practical</i>
<i>practical</i>	<i>active,constructive,economical,sophisticated,discriminating</i>
<i>discriminating</i>	<i>aesthetic,eclectic,imaginative,practical,rational,polished</i>
<i>rational</i>	<i>advisable,analytical,wise,discriminating,intelligent,logical</i>
<i>logical</i>	<i>analytical,intelligent,rational</i>
<i>analytical</i>	<i>logical,rational</i>
<i>wise</i>	<i>advisable,rational,intelligent</i>
<i>advisable</i>	<i>rational,wise</i>
<i>mechanical</i>	<i>instinctive,programmed,standardized</i>
<i>imaginative</i>	<i>Discriminating</i>
<i>aesthetic</i>	<i>discriminating,polished</i>
<i>polished</i>	<i>aesthetic,discriminating,hospitable,informative,social,poised</i>
<i>social</i>	<i>amiable,cordial,hospitable,informative,sociable,affable,convivial,friendly,genial,gregarious,polished,political</i>
<i>sociable</i>	<i>cooperative,friendly,social,affable,genial,gregarious</i>

<i>cooperative</i>	<i>coherent,sociable</i>
<i>coherent</i>	<i>Cooperative</i>
<i>gregarious</i>	<i>convivial,cosmopolitan,sociable,social,friendly,genial</i>
<i>cosmopolitan</i>	<i>impartial,gregarious,sophisticated</i>
<i>sophisticated</i>	<i>cosmopolitan,practical</i>
<i>programmed</i>	<i>Mechanical</i>
<i>impartial</i>	<i>Cosmopolitan</i>
<i>convivial</i>	<i>gregarious,social,friendly,genial</i>
<i>genial</i>	<i>amiable,convivial,cordial,gregarious,hearty,sociable,social,cheerful</i>
<i>cheerful</i>	<i>genial,hearty,resilient,optimistic</i>
<i>resilient</i>	<i>Cheerful</i>
<i>hearty</i>	<i>cheerful,cordial,devout,fresh,genial,strong,enthusiastic,generous</i>
<i>strong</i>	<i>alcoholic,energetic,fresh,hard,hearty,"in shape",courageous,influential</i>
<i>in shape</i>	<i>Strong</i>
<i>alcoholic</i>	<i>hard,strong</i>
<i>hard</i>	<i>alcoholic,severe,austere,inflexible,shrewd,strong</i>
<i>standardized</i>	<i>Mechanical</i>
<i>severe</i>	<i>ambitious,ascetic,austere,exact,hard,serious,inflexible</i>
<i>exact</i>	<i>Severe</i>
<i>ambitious</i>	<i>energetic,enterprising,industrious,severe</i>

<i>energetic</i>	<i>ambitious, strong, enterprising</i>
<i>enterprising</i>	<i>ambitious, energetic</i>
<i>industrious</i>	<i>Ambitious</i>
<i>ascetic</i>	<i>severe, reclusive</i>
<i>reclusive</i>	<i>ascetic, reserved</i>
<i>reserved</i>	<i>irritable, reclusive, inhibited</i>
<i>irritable</i>	<i>excitable, sensitive, sour, reserved</i>
<i>intuitive</i>	<i>Instinctive</i>
<i>sour</i>	<i>austere, irritable</i>
<i>austere</i>	<i>hard, severe, formal, serious, sour</i>
<i>serious</i>	<i>austere, dignified, severe</i>
<i>dignified</i>	<i>elevated, formal, serious</i>
<i>formal</i>	<i>austere, dignified, conventional</i>
<i>conventional</i>	<i>Formal</i>
<i>elevated</i>	<i>dignified, eloquent</i>
<i>eloquent</i>	<i>Elevated</i>
<i>sensitive</i>	<i>Irritable</i>
<i>inhibited</i>	<i>Reserved</i>
<i>emotional</i>	<i>sentimental, impulsive</i>
<i>inflexible</i>	<i>hard, severe</i>
<i>courageous</i>	<i>Strong</i>

<i>influential</i>	<i>Strong</i>
<i>fresh</i>	<i>original,active,alert,hearty,strong</i>
<i>original</i>	<i>Fresh</i>
<i>generous</i>	<i>Hearty</i>
<i>cordial</i>	<i>genial,friendly,hearty,social</i>
<i>friendly</i>	<i>convivial,cordial,gregarious,kind,social,affectionate,sociable</i>
<i>affectionate</i>	<i>friendly,kind-hearted,responsive</i>
<i>responsive</i>	<i>affectionate,kind-hearted</i>

A lista de adjacência apresentada acima se refere ao grafo correspondente às ligações entre 222 traços, tal como descritas no True Term Thesaurus Database. Estes traços serão utilizados nas aplicações computacionais dos capítulos 6 e 7.

Esta lista de adjacência só contém termos cujos vértices associados no mencionado grafo possuem ligações com pelo menos outro vértice. Vértices que ficaram totalmente isolados no grafo não estão incluídos nessa lista de adjacência.

Na lista de adjacência acima, para cada palavra na coluna da esquerda têm-se todas as palavras relacionadas (em algum sentido) a ela, no TTT.

5.5.5.3 Rotinas de geração de matrizes

As rotinas em questão são responsáveis pela geração dos relatórios:

- Matrizes de adjacência do grafo. Estas matrizes representam, por meio de uma matriz esparsa, as relações de adjacência ou não, entre cada par de palavras do banco.

- Matrizes de Markov (que é *input* para o algoritmo MCL, de clusterização em grafos).

5.5.5.4 Rotinas relacionadas ao algoritmo MCL e *Ratings* de *Clusters* encontrados pelo MCL

5.5.5.4.1 Rotinas relativas às funcionalidades do algoritmo MCL (Markov Clustering).

Estas rotinas geram os seguintes relatórios:

- Relatórios de *clusters* encontrados pelo algoritmo MCL. Este relatório lista cada *cluster* identificado e os vértices do grafo pertencentes a cada um destes agrupamentos. Um exemplo deste relatório se encontra no Apêndice deste trabalho, com aplicações em traços de personalidade.

-Relatório contendo a matriz de resultado final do algoritmo MCL. Este relatório se encontra no DVD (*Digital Versatile Disk*) anexo da tese.

- Relatório contendo todas as matrizes, ao longo de cada iteração do algoritmo. Este relatório também se encontra no DVD anexo.

- Questionário em arquivo. xls. Este relatório visa aplicações nas áreas organizacionais e clínicas envolvendo traços de personalidade presentes no thesaurus empregado. Este relatório é escrito em arquivo. xls e contém os traços selecionados para uma dada aplicação (e seus antônimos) de cada *cluster*, assim como campos para que *ratings* sejam dados em relação a cada traço. Um exemplo deste relatório gerado (e preenchido) se encontra no Apêndice.

5.5.6 Rotinas geradoras de listas de adjacência

Tais rotinas geram listas de adjacência do grafo subjacente às relações entre palavras presentes em uma importação do *thesaurus* no banco de dados.

5.6 HARDWARES UTILIZADOS

Para as implementações computacionais realizadas neste trabalho, várias arquiteturas diferentes foram utilizadas. Boa parte dos algoritmos desenvolvidos em Java e no sistema em php foi rodada em um computador Dell i7 com CPU de 3.7 GHz, 8GB RAM, e sistema operacional Windows 7 Professional 64 bits.

Algumas implementações foram implementadas no serviço de *cloud computing* da Amazon, denominado *Amazon Web Services*. Na maior parte dos casos em que este serviço foi utilizado foi empregado o seguinte *hardware*: CPU Intel Xeon E5- 26650 2.4 GHz com 34.2 GB de RAM e sistema operacional Windows Server 2012.

Foram utilizados também duas arquiteturas para implementações computacionais de teste (com pequenas quantidades de dados) e para edição de textos. São elas: um notebook Dell (Vostro 1500) com CPU Core 2 Duo de 2.0 GHz e 2,5 GB de RAM e um *desktop* com CPU Intel P4 3.6 GHz com 2 GB RAM e Windows XP *Home edition*. Ambas as plataformas 32 bits.

5.7 RELAÇÃO ENTRE RELATÓRIOS GERADOS, ALGORITMOS DE CLUSTERIZAÇÃO E CTPT'S

Com o objetivo de tornar mais clara a forma como alguns dos relatórios gerados pelo sistema PHP e pelas rotinas em Java se relacionam entre si e com os algoritmos de clusterização (com o objetivo final de se determinar CTPT's) a figura 32 abaixo apresenta um diagrama em blocos.

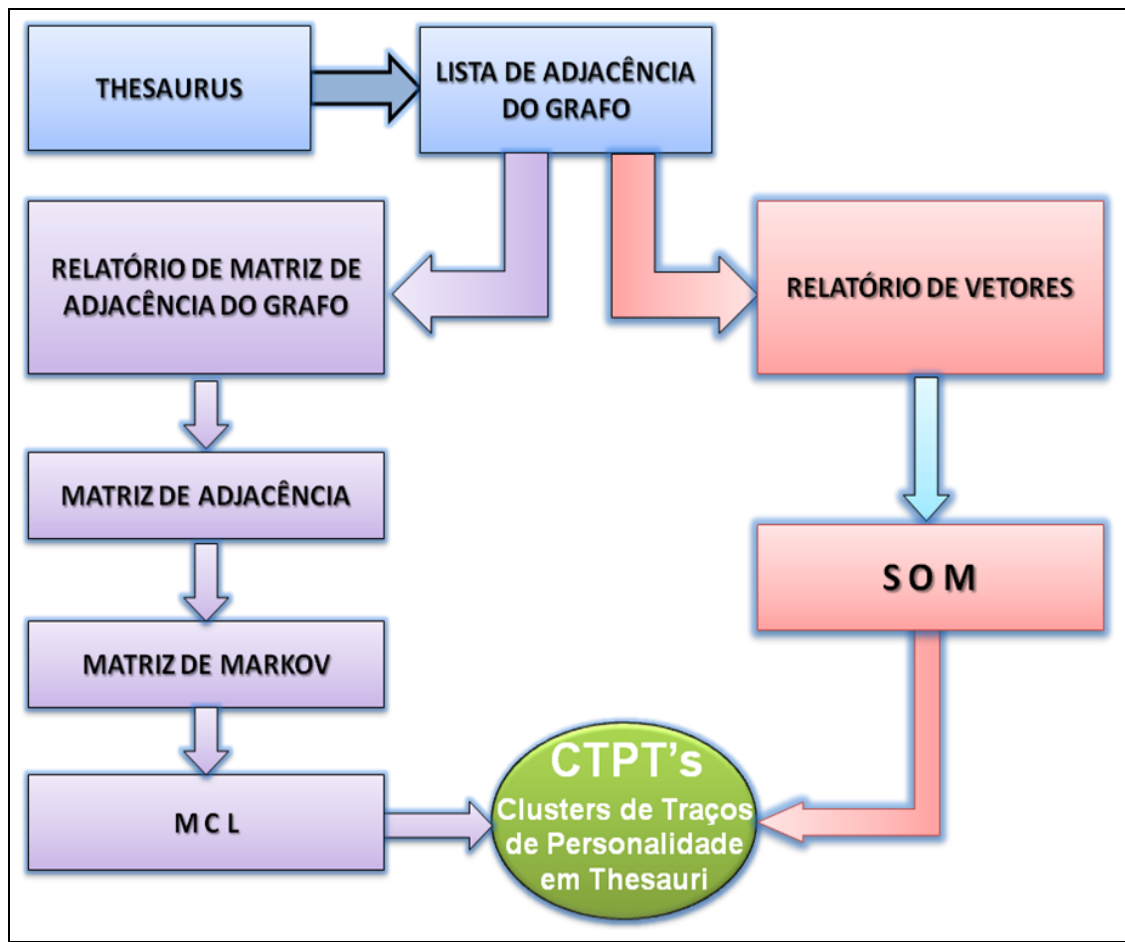


Figura 32: Relação entre relatórios gerados, pelo Sistema em PHP e pelas rotinas em Java e os algoritmos de clusterização utilizados com o objetivo de se determinar CTPT's.

Pela análise da figura 32 acima, verifica-se que a lista de adjacência gerada tanto pelo sistema PHP como pelas rotinas em Java é o relatório-base, a partir do qual os demais serão gerados.

Seguindo-se a direção das setas na figura supracitada, tem-se a informação de qual relatório serve de *input* para outra rotina que gerará um outro relatório a partir do primeiro.

Deste modo, a lista de adjacência gerada sobre um grafo é o *input* necessário tanto para a geração da matriz de adjacência do grafo (efetuado tanto pelo sistema PHP como pelas rotinas em Java) como para a geração do relatório de vetores (feita somente no sistema PHP).

A partir da matriz de adjacência pode-se gerar matrizes de Markov, que serão a entrada para o algoritmo MCL. Esta geração desta última matriz a partir da primeira é efetuada por rotina específica em linguagem Java.

O relatório de vetores é a entrada final a ser utilizado pelas redes neurais SOM.

Ambos os algoritmos de clusterização geram então CTPT's a partir dos *clusters* identificados no grafo representado a rede de ligações entre os traços considerados em uma determinada aplicação, tal como presentes do *True Term Thesaurus Database*.

CAPÍTULO 6: DETALHES DE IMPLEMENTAÇÃO COM DIFERENTES ALGORITMOS DE CLUSTERIZAÇÃO

O objetivo deste capítulo é apresentar os detalhes de implementação e os motivos da realização de um estudo contendo poucos (mas extremamente importantes na literatura de psicologia) adjetivos-traços de personalidade no *thesaurus* utilizado, os detalhes de implementação e os resultados da aplicação desses métodos quantitativos envolvendo essa pequena quantidade de palavras do *thesaurus*.

Com isso, o objetivo principal aqui é se ganhar *insights* sobre as metodologias e particularidades de cada uma delas quando lidarem com diferentes (e maiores) conjuntos de dados.

Esta etapa também visa a reconhecer potenciais problemas não antecipados com cada metodologia de clusterização adotada e comparar seus resultados.

Com isto, pretende-se determinar qual a metodologia mais adequada para a posterior determinação de agrupamentos de palavras em outros conjuntos de dados do mesmo TTT.

A escolha do algoritmo deverá levar em conta os resultados obtidos, dificuldades de implementação e demanda computacional de cada um.

Para tais propósitos, partiu-se de um conjunto de 222 palavras. Este conjunto é formado em boa parte pelas 171 palavras (variáveis) sugeridas por CATTELL (1943b). As demais palavras são substitutas próximas em sentido das palavras sugeridas por Cattell não presentes na base de dados *Thesaurus*.

O que se deseja principalmente, neste sentido, é a determinação de clusters de palavras próximas em sentido, em um pequeno conjunto de palavras. Desejavelmente, estes clusters não devem se sobrepor uns aos outros. O motivo de sobreposições de clusters não serem desejadas são:

A existência de *clusters* não sobrepostos (isto é, com interseção vazia) implica em maior facilidade na seleção de palavras que melhor representem os sentidos mais gerais de cada cluster.

Clusters não sobrepostos eliminam os problemas de se ter que arbitrar sobre a partir de que tamanho as interseções entre eles deveriam ou não ser consideradas como *clusters* isolados.

Finalmente procede-se a uma comparação entre os agrupamentos encontrados pelas diferentes metodologias com os *clusters* finais obtidos por Catell após agrupamento dos

resultados de aplicações de questionários (contendo gradações entre as 171 palavras por ele apresentadas e os seus antônimos) por meio de técnicas de correlação.

Embora se possa levantar (relevantemente) a questão de que Cattell tenha chegado nestes clusters finais baseado nas respostas de questionários, e que, portanto, os agrupamentos por ele encontrados representam traços que estão próximos nos indivíduos (o que é diferente do que se está fazendo aqui, onde se está determinando clusters de traços próximos em sentido; de acordo com a estrutura de ligações entre elas, no TTT) algumas observações interessantes podem ser feitas nessas comparações.

Ademais, o que se pretende no próximo capítulo, é apresentar uma forma sistemática de agrupamento de palavras (e logo também de adjetivos-traços) por proximidade de sentido; onde, a partir desses clusters se possa mais facilmente agrupar os traços em categorias de proximidade de aspectos da personalidade sem se ter que fazer a leitura exaustiva de um *thesaurus* inteiro.

Lembrar também na tese de que estamos cientes de que as associações entre palavras no TTT podem, em alguns casos não serem de imediato reconhecimento a um determinado leitor; mas nestes casos isto se deve ao fato de que são essas as ligações presentes no TTT. Isto é, reconhecemos a possibilidade de que algumas associações entre palavras presentes no repositório em questão podem não ser imediatas (ou mesmo aceitas) por especialistas da psicologia e demais áreas comportamentais; mas este é um pequeno risco enfrentado para que se possam aplicar métodos sistemáticos na busca de clusters de palavras relacionadas (e, portanto se evitar o método de Cattell, da leitura e coleta exaustivas de palavras no TTT).

Como já exposto anteriormente, apesar dos *clusters* encontrados pela rede de Kohonen conterem palavras relacionadas mais proximamente entre si, de acordo com o a linguagem natural, esse reconhecimento dessa proximidade pode nem sempre se dar de forma unânime entre diferentes pessoas, em função da ambiguidade que os sentidos das palavras podem ganhar em diferentes culturas e também em função de diferenças de domínio de vocabulário.

6.1 CARACTERÍSTICAS DO CONJUNTO DE PALAVRAS SELECIONADAS

Como mencionado anteriormente, parte das 222 palavras utilizadas neste estudo corresponde exatamente a alguns dos 171 traços do estudo de Cattell e o restante as substitutas das demais palavras (do conjunto das 171 apresentadas por Cattell) presentes no banco *Thesaurus* usado.

Estas 171 palavras representam um conjunto de traços a partir do qual diversos modelos de grandes fatores da personalidade foram desenvolvidos; assim, este pequeno conjunto de palavras tem grande capacidade de descrição da personalidade humana.

Contudo, a tarefa de se encontrar um pequeno conjunto de traços com estas propriedades não é uma tarefa fácil. De acordo com Cattell (1943b, p. 491).

A persistent difficulty in this task of reducing any systematizing personality trait terms arose from the great variability of width or area of meaning of the different terms. While most described single traits, a minority really had reference to "types" and might be described as syndrome or "blanket" terms, e.g., old-maidish, Mephistophelean, neurotic, having the qualities of a leader, extravert, gentlemanly, and various occupational syndromes. These very wide terms, like some exceptionally narrow ones mentioned above, were set aside for special handling in connection with various additions to the Allport-Odbert dictionary list".

A tabela a seguir apresenta estas 222 palavras selecionadas.

Tabela 22: As 222 palavras selecionadas para a aplicação das metodologias de clusterização.

Numeração das 171 palavras propostas por Cattell	Palavra Original proposta por Cattell	Numeração das 222 palavras escolhidas	Palavra Candidata a Substituta da Original	Tradução da palavra	Fonte da verificação da associação entre a palavra original e a substituta proposta	Antônimos	Tem ou não no <i>thesaurus</i> utilizado	Tradução Antônimo
1	<i>Abilities: dextery</i>	1	<i>skilful</i>	habilidoso	Thesaurus TT software	<i>unskillful</i>	sim	inapto
1	<i>Abilities: dextery</i>	2	<i>deft</i>	destro, hábil	dicionário Webster	<i>inept</i>	sim	inapto
1	<i>Abilities: dextery</i>	3	<i>skilled</i>	hábil	Thesaurus TT software	<i>unskilled</i>	sim	sem habilidade
1	<i>Abilities: dextery</i>	4	<i>proficient</i>	proficiente	Thesaurus TT software	<i>incapable</i>	sim	incapaz
1	<i>Abilities: dextery</i>	5	<i>dexterous</i>	destro, hábil com as mãos	Thesaurus TT software	<i>unskilled</i>	sim	sem habilidade
2	<i>Abilities: drawing</i>	6	<i>pictorial</i>	pictórico, visual	Thesaurus TT software	<i>not pictorial/visual/descriptive</i>	não	não-pictórico/visual/descriptivo
3	<i>Abilities: intelligence</i>	7	<i>intelligent</i>	inteligente	Thesaurus TT software	<i>unintelligent, stupid</i>	sim, ambas	não-inteligente, estúpido, burro
4	<i>Abilities: mathematical</i>	8	<i>exact</i>	exato, preciso, matemático	Thesaurus TT software	<i>inexact</i>	sim	inexato, impreciso, não-matemático
5	<i>Abilities: mechanical aptitude</i>	9	<i>instinctive</i>	instintivo	Thesaurus TT software	<i>learned, conscious</i>	Sim, todas	culto, aprendido, consciente

5	<i>Abilities: mechanical aptitude</i>	10	<i>mechanical</i>	mecânico, automatizado, quem faz as coisas automaticamente	Thesaurus TT software	<i>not mechanical, planned</i>	não	não mecânico, que não faz as coisas automaticame nte, sem pensar
5	<i>Abilities: mechanical aptitude</i>	11	<i>programme d</i>	mecânico, automatizado, quem faz as coisas automaticamente	Thesaurus TT software	<i>not programmed, planned</i>	não	não mecânico, de forma programada e automática.
5	<i>Abilities: mechanical aptitude</i>	12	<i>standardize d</i>	mecânico, automatizado, quem faz as coisas automaticamente	Thesaurus TT software	<i>not standarized, planned</i>	não	não- mecânico, não- padronizado, não- planejado.
5	<i>Abilities: mechanical aptitude</i>	13	<i>perfunctory</i>	superficial, perfunctório, desinteressado. Sem entusiasmo, sem pensar	Thesaurus TT software	<i>not perfunctory</i>	não	não- automático/se m pensar /sem interesse.
6 e 7	<i>Abilities: musical aptitude E Interests special: music</i>	14	<i>musical</i>	musical (interesse e habilidades)	Thesaurus TT software	<i>not musical</i>	não	não-musical
8	<i>Abilities: physical endurance</i>	15	<i>in shape</i>	em forma, com resistência física	Thesaurus TT software	<i>in bad shape</i>	sim	fora de forma, machucado, doente, sem resistência física
8	<i>Abilities: physical endurance</i>	16	<i>physical</i>	físico	Thesaurus TT software	<i>in bad shape</i>	sim	fora de forma, machucado, doente, sem resistência física
8	<i>Abilities: physical endurance</i>	17	<i>corporeal</i>	corporeal	Thesaurus TT software	<i>in bad shape</i>	sim	fora de forma, machucado, doente, sem resistência física
8	<i>Abilities: physical endurance</i>	18	<i>animal</i>	animal	Thesaurus TT software	<i>in bad shape</i>	sim	fora de forma, machucado, doente, sem resistência física
8	<i>Abilities: physical endurance</i>	19	<i>bodily</i>	encorpado	Thesaurus TT software	<i>in bad shape</i>	sim	fora de forma, machucado, doente, sem resistência física

8	<i>Abilities: physical endurance</i>	20	<i>corporal</i>	corporal	Thesaurus TT software	<i>in bad shape</i>	sim	fora de forma, machucado, doente, sem resistência física
9	<i>abilities: reasoning</i>	21	<i>rational</i>	racional	Thesaurus TT software	<i>irrational</i>	sim	irracional, ilógico
9	<i>abilities: reasoning</i>	22	<i>logical</i>	lógico	Thesaurus TT software	<i>illogical</i>	sim	irracional, ilógico
10	<i>abilities: spatial, visual ability</i>	23	<i>visionary</i>	visionário	Thesaurus TT software	<i>not visionary</i>	não	não-visionário
10	<i>abilities: spatial, visual ability</i>	24	<i>visible</i>	visível	Thesaurus TT software	<i>discrete</i>	sim	discreto
10	<i>abilities: spatial, visual ability</i>	25	<i>visual</i>	visual	Thesaurus TT software	<i>not visual</i>	não	não-visual
10	<i>abilities: spatial, visual ability</i>	26	<i>observing</i>	observante, que observa, atento, alerta	Thesaurus TT software	<i>blunt</i>	sim	cego
10	<i>abilities: spatial, visual ability</i>	27	<i>observer</i>	observador	Thesaurus TT software	<i>blunt</i>	sim	cego
11	<i>Abilities: verbal attitude</i>	28	<i>lexical</i>	léxico	Thesaurus TT software	<i>written, literal</i>	sim	literal, quieto
11	<i>Abilities: verbal attitude</i>	29	<i>verbal</i>	verbal	Thesaurus TT software	<i>written, literal</i>	sim	literal, quieto
11	<i>Abilities: verbal attitude</i>	30	<i>vocal</i>	vocal	Thesaurus TT software	<i>written, literal</i>	sim	literal, quieto
11	<i>Abilities: verbal attitude</i>	31	<i>speaking</i>	falante	Thesaurus TT software	<i>quiet</i>	sim	literal, quieto
11	<i>Abilities: verbal attitude</i>	32	<i>unwritten</i>	verbal, não-escrito, de boca	Thesaurus TT software	<i>written, literal</i>	sim	literal, quieto
11	<i>Abilities: verbal attitude</i>	33	<i>open-hearted</i>	aberto, de coração aberto, franco	Thesaurus TT software	<i>written, literal</i>	sim	literal, quieto
12	<i>acquisitive</i>	34	<i>acquisitive</i>	ávido, ganancioso, aquisitivo	-	<i>non acquisitive</i>	não	não-ávido, não-ganancioso
13	<i>affected</i>	35	<i>affected</i>	aflito, afetado	-	<i>natural (Catell), untroubled</i>	sim, ambas	natural, sem problemas
14	<i>affectionate</i>	36	<i>affectionate</i>	que demonstra amor, carinhoso	-	<i>uncaring</i>	não	quem não liga
15	<i>agoraphobic</i>	37	<i>frightened</i>	espantado, aterrorizado	Thesaurus TT software e WordNet 3.0	<i>bold</i>	sim	corajoso, confiante
15	<i>agoraphobic</i>	38	<i>panicky</i>	em pânico, assustado	Thesaurus TT software e WordNet 3.0	<i>unafraid</i>	sim	sem medo

15	<i>agoraphobic</i>	39	<i>coward</i>	covarde	Thesaurus TT software e WordNet 3.0	<i>brave</i>	sim	bravo, destemido
15	<i>agoraphobic</i>	40	<i>reclusive</i>	recluso, reclusivo	Thesaurus TT software e WordNet 3.0	<i>gregarious, sociable</i>	sim	sociável
16	<i>alcoholic</i>	41	<i>alcoholic</i>	alcólicico, contagiante	-	<i>non-alcoholic</i>	não	não-contagiante
17	<i>alert</i>	42	<i>alert</i>	alerta	-	<i>distracted</i>	sim	distraído, desatento
18	<i>ambitious</i>	43	<i>ambitious</i>	ambicioso	dicionário Webster	<i>unambitious</i>	não	não-ambicioso
19	<i>amorous</i>	44	<i>amorous</i>	romântico, ardente	-	<i>lustless*</i>	não	lângido, destituído de desejo sexual
20	<i>analytical</i>	45	<i>analytical</i>	analítico, bom em análises	-	<i>synthetical, unsystematic, cursory</i>	não	sintético, não-sistemático, superficial e apressado.
21	<i>antever</i>	46	<i>preventive</i>	preventivo	Webster Online e dicionário Webster	<i>remedial</i>	sim	corretivo, remedial
22	<i>argumentative</i>	47	<i>argumentative</i>	argumentativo, contestador	-	<i>agreeable</i>	sim	quem concorda, concorde
23	<i>arrogant</i>	48	<i>arrogant</i>	arrogante	-	<i>humble*</i>	sim	humilde
24	<i>ascetic</i>	49	<i>ascetic</i>	disciplinado, abstinente,	-	<i>sensuous*</i>	sim	voluptoso, libidinoso, que apela aos sentidos, sensual
25	<i>assertive</i>	50	<i>assertive</i>	assertivo, decisivo	-	<i>submissive*, passive</i>	sim, todas	submisso, passivo
26	<i>austere</i>	51	<i>austere</i>	rígido, austero, severo, sério	-	<i>flexible, gentle</i>	sim	flexível, gentil
27	<i>autocratic</i>	52	<i>autocratic</i>	autocrático, despótico	-	<i>democratic</i>	sim	democrático
28	<i>boastful</i>	53	<i>boastful</i>	pretencioso, presunsoço	-	<i>modest*</i>	sim	modesto
29	<i>brooding</i>	54	<i>complaining</i>	que reclama	dicionário Webster	<i>unrepining (Catell), enjoying</i>	não	quem não reclama
30	<i>cautious</i>	55	<i>cautious</i>	cauteloso	-	<i>reckless</i>		descuidado, despreocupado
31	<i>charming</i>	56	<i>charming</i>	charmoso, cativante	-	<i>repulsive</i>	sim	repulsivo
32	<i>cheerful</i>	57	<i>cheerful</i>	alegre, animado	-	<i>gloomy*</i>	sim	triste, melancólico
33	<i>claustrophobic</i>	58	<i>afraid</i>	com medo	dicionário Webster	<i>unafraid</i>	sim	sem medo
34	<i>clear thinking</i>	59	<i>coherent</i>	coerente	dicionário Webster	<i>incoherent*</i>	sim	incoerente
35	<i>clever</i>	60	<i>clever</i>	esperto, "safo", perspicaz	-	<i>unclever, clumsy, awkward</i>	sim as duas últimas	desajeitado, sem habilidade

36	<i>conceited</i>	61	<i>conceited</i>	vaidoso, convencido, satisfeito consigo mesmo	-	<i>dissatisfied*</i>	não	insatisfeito
37	<i>conscientious</i>	62	<i>conscientious</i>	consciente, escrupuloso	-	<i>conscienceless (Catell), not diligent, careless</i>	sim, as duas últimas	sem consciência, sem diligência, descuidado.
38	<i>constructive</i>	63	<i>constructive</i>	construtivo	-	<i>destructive</i>	sim	destrutivo
39	<i>contented</i>	64	<i>contented</i>	contente	-	<i>dissatisfied (Catell)</i>	sim	insatisfeito, descontente
40	<i>conventional</i>	65	<i>conventional</i>	convencional	-	<i>individualistic (Catell), unique</i>	o segundo sim	único
41	<i>cooperative</i>	66	<i>cooperative</i>	cooperativo	-	<i>obstructive</i>	sim	obstrutivo
42	<i>courageous</i>	67	<i>courageous</i>	corajoso	-	<i>cowardly</i>	sim	covarde
43	<i>curious</i>	68	<i>curious</i>	curioso	-	<i>unenquiring, incurious, indifferent</i>	sim, a última	não-curioso, indiferente
44	<i>cynical</i>	69	<i>cynical (Catell), unbelieving, disingenuous</i>	cínico, pessoa que não acredita em nada	-	<i>idealistic</i>	sim	idealista
45	<i>debonnaire</i>	70	<i>cosmopolitan</i>	cosmopolita, multicultural, pessoa viajada	dicionário Webster	<i>not cosmopolitan</i>	não	não-cosmopolitano
46	<i>defensive</i>	71	<i>closed</i>	fechado	Thesaurus TT software	<i>open</i>	sim	aberto, franco
47	<i>dubitative</i>	72	<i>doubtful</i>	em dúvida	dicionário Webster	<i>decisive</i>	sim	decisivo
48	<i>easy going</i>	73	<i>easygoing</i>	fácil de lidar	dicionário Webster	<i>short-tempered (Catell)</i>	sim	pavio curto
49	<i>eccentric</i>	74	<i>eccentric</i>	excêntrico	-	<i>inconspicuous</i>	sim	imperceptível, discreto
50	<i>effeminate</i>	75	<i>effeminate</i>	efeminado	-	<i>masculine (Catell), manly</i>	sim	másculo, robusto
51	<i>egotistical (Catell), self-centered</i>	76	<i>egotistical</i>	egoístico	-	<i>altocentric*</i>	não	não-centrado em si mesmo
52	<i>eloquent</i>	77	<i>eloquent</i>	eloquente, expressivo	-	<i>inarticulate</i>	sim	não-eloquente, inarticulado
53	<i>emotional I(all varieties)</i>	78	<i>emotional</i>	emocional, emotivo	Thesaurus TT software	<i>unemotional*</i>	sim	não-emotivo
54 b	<i>emotional II (sociability)</i>	79	<i>convivial</i>	social, convivente	Dicionário Webster e Thesaurus TT software	<i>Unemotional (Catell), lethargic</i>	sim a última	não-convivente
54 b	<i>emotional II (sociability)</i>	80	<i>affable</i>	afável	Thesaurus TT software	<i>unmannerly</i>	sim	sem modos, desrespeitoso
55	<i>energetic industrious</i>	81	<i>industrious</i>	industrioso	dicionário Webster	<i>indolent*</i>	sim	indolente

56	<i>energetic-spirited</i>	82	<i>energetic</i>	energético	dicionário Webster	<i>languid*</i>	sim	lemto em seus movimentos, lângido, letárgico
57	<i>enterprising</i>	83	<i>enterprising (Catell), resourceful</i>	empreendedor	dicionário Webster	<i>shiftless (Catell), unadventurous</i>	sim, ambas	não-providente, não-emprededor, não-aventuroso
58	<i>enthusiastic</i>	84	<i>enthusiastic</i>	entusiasmado	-	<i>apathetic*</i>	Sim	apático
59	<i>evasive</i>	85	<i>evasive (Catell), avoiding, evading</i>	evasivo	-	<i>facing Life (Catell)</i>	não	enfrenta a vida
60	<i>excitable</i>	86	<i>excitable</i>	excitável	-	<i>phlegmatic (Catell), unexcitable</i>	não	não-excitável
61	<i>exhibitionistic</i>	87	<i>perceivable</i>	percebível, exibicionista	Webster Online e dicionário Webster	<i>self-effacing (Catell), reticent</i>	sim, a segunda	reticente
62	<i>extra-punitive</i>	88	<i>severe</i>	severo, punitivo	Thesaurus TT software	<i>praiseful (Catell), appreciative</i>	sim, a segunda	apreciativo
63	<i>fair-minded</i>	89	<i>impartial</i>	imparcial	dicionário Webster	<i>partial (Catell), unfair</i>	sim	parcial, injusto
64	<i>fastidious</i>	90	<i>fastidious (Catell), meticulous</i>	meticuloso	-	<i>coarse (Catell), uncritical, unrefined, unfussy</i>	sim, as três primeiras	não-meticuloso, não-crítico, não-refinado (quem não refina as coisas).
65	<i>flattering</i>	91	<i>flattering</i>	lisonjeiro, elogioso	-	<i>unflattering</i>	não	não-lisonjeiro
66	<i>formal</i>	92	<i>formal</i>	formal	-	<i>casual (Catell), informal</i>	sim, ambas	informal
67	<i>frank</i>	93	<i>frank</i>	franco	-	<i>secretive*</i>	sim	secretivo
68	<i>friendly</i>	94	<i>friendly</i>	amigável	-	<i>hostile (Catell), unfriendly</i>	sim, ambas	não-amigável, hostil
69	<i>generous</i>	95	<i>generous</i>	generoso	-	<i>tight-fisted*</i>	sim	mão-fechada, não-generoso
70	<i>genial</i>	96	<i>genial</i>	cordial, afável	-	<i>cold-hearted (Catell), rude</i>	sim, a última	rude, não-gentil, não-afável
71	<i>gluttonous</i>	97	<i>gluttonous</i>	voraz, glutão	-	<i>queasy*</i>	sim	enjoado, embrulhado
72	<i>grateful</i>	98	<i>grateful</i>	grato	-	<i>thankless*</i>	sim	ingrato
73	<i>habit-bound</i>	99	<i>habituated</i>	habituaado, acostumado	dicionário Webster	<i>labile (Catell), fleeting, transient</i>	sim as duas últimas	transiente, transitório, que muda sempre tudo

74	<i>hard</i>	100	<i>hard</i>	duro, amargo	-	<i>soft-hearted*</i>	sim	coração-mole, benevolente, com piedade
75	<i>headstrong</i>	101	<i>headstrong</i>	cabeçudo, obstinado, teimoso	-	<i>gentle tempered (Catell), not stubborn</i>	sim, a última	não-teimoso
76	<i>hearty</i>	102	<i>hearty</i>	forte, cordial, bonachão, substancioso	-	<i>quiet*</i>	sim	quieto
77	<i>high-strung</i>	103	<i>high-strung</i>	excitável, impaciente	-	<i>relaxed*</i>	sim	relaxado
78	<i>honest</i>	104	<i>honest</i>	honesto	-	<i>dishonest*</i>	sim	desonesto
79	<i>hurried</i>	105	<i>hurried</i>	apressado	-	<i>lethargic*</i>	sim	letárgico
80 a	<i>hypochondriacal (Catell), hypochondriac</i>	106	<i>anxious</i>	ansioso	dicionário Webster	<i>calm</i>	sim	calmo
80 b	<i>hypochondriacal (Catell), neurotic, hypochondriac</i>	158	<i>neurotic</i>	neurótico	WordNet 3.0 e Thesaurus TT software	<i>not neurotic</i>	não	não-neurótico
80 c	<i>hypochondriacal (Catell), hypochondriac</i>	107	<i>splenetic</i>	irritável, petulante	WordNet 3.0 e Thesaurus TT software e thesaurus.com	<i>happy</i>	sim	feliz
81	<i>imaginative</i>	108	<i>imaginative</i>	imaginativo	-	<i>dull*</i>	sim	nublado, tedioso, estupificado
82	<i>imitative</i>	109	<i>imitative</i>	imitativo	-	<i>genuine</i>	sim	genuíno
83	<i>impulsive temperamentally</i>	110	<i>impulsive</i>	impulsivo	Thesaurus TT software	<i>deliberate (Catell), planned</i>	Sim, as duas últimas	planejado
84	<i>independent</i>	111	<i>independent</i>	independente	-	<i>dependent</i>	sim	dependente
85	<i>inflexible (emotionally)</i>	112	<i>inflexible</i>	inflexível	Thesaurus TT software	<i>adaptable</i>	sim	adaptável
86	<i>inhibited</i>	113	<i>inhibited</i>	inibido	-	<i>incontinent (Catell), outgoing</i>	sim, a segunda	extrovertido
87	<i>Interests special: aesthetic (general)</i>	114	<i>aesthetic</i>	interessado em estética, estético, culto, artístico	Thesaurus TT software	<i>not aesthetic</i>	não	não estético/culto/artístico
88	<i>Interests special: artistic</i>	115	<i>decorative</i>	decorativo, estiloso	Thesaurus TT software	<i>plain</i>	sim	simples, comum,
88	<i>Interests Special: artistic</i>	116	<i>discriminating</i>	inventivo, refinado, educado, com bom gosto	Thesaurus TT software	<i>plain</i>	sim	simples, comum,

89	<i>Interests special: economic</i>	117	<i>economical</i>	econômico, interessado em economia	Thesaurus TT software	<i>not economic, not interested in economics</i>	não	não econômico, não interessado em Economia
89	<i>Interests Special: economic</i>	118	<i>frugal</i>	frugal	Thesaurus TT software	<i>wasteful</i>	sim	esbanjador, perdulário, desperdiçador
89	<i>Interests Special: economic</i>	119	<i>provident</i>	providente, quem provê	Thesaurus TT software	<i>improvident</i>	sim	improvidente, quem não provê
89	<i>Interests Special: economic</i>	120	<i>thrifty</i>	frugal, sóbrio	Thesaurus TT software	<i>spendthrift</i>	sim	palyboy, gastador
89	<i>Interests Special: economic</i>	121	<i>sparing</i>	modesto, frugal	Thesaurus TT software	<i>generous, careless</i>	sim	generoso, descuidado
90	<i>Interests special: home & family</i>	122	<i>devout</i>	devoto, dedicado, devotado, ortodoxo, aderente, conservado	Webster Online, Merriam-Webster Online, WordNet 2.1 Browser e TT software Thesaurus, acessos de 17-01-12 a 20-01-2012	<i>not devout</i>	não	não-devoto/ortodoxo, aderente/conservador
91	<i>Interests special: physical activity</i>	123	<i>sportive</i>	esportivo	Thesaurus TT software	<i>not sportive</i>	não	não-esportivo.
92	<i>Interests special: political</i>	124	<i>political</i>	político (capacidade de transitar entre os diferentes pontos de vistas e posições políticas)	Thesaurus TT software	<i>not-political</i>	não	não-político
93	<i>Interests special: religious</i>	125	<i>religious</i>	religioso	Thesaurus TT software	<i>irreligious</i>	não	não-religioso
94	<i>Interests special: social</i>	126	<i>social</i>	social, convivente	Thesaurus TT software	<i>antisocial</i>	sim	anti-social
94	<i>Interests special: social</i>	127	<i>informative</i>	informativo	Thesaurus TT software	<i>uninformative</i>	não	não-informativo
94	<i>Interests special: social</i>	128	<i>hospitable</i>	hospitaleiro	Thesaurus TT software	<i>not hospitable</i>	não	não-hospitaleiro
94	<i>Interests special: social</i>	129	<i>amiable</i>	amável, afável	Thesaurus TT software	<i>hostile</i>	sim	hostil
94	<i>Interests special: social</i>	130	<i>cordial</i>	cordial	Thesaurus TT software	<i>disagreeable</i>	sim	desagradável
95	<i>Interests special: technical</i>	131	<i>professional</i>	profissional, técnico	Thesaurus TT software	<i>unprofessional</i>	sim	profissional, técnico

96	<i>Interests special: theoretical</i>	132	<i>abstract</i>	abstrato, teórico	Thesaurus TT software, WordReference.com	<i>concrete, simple</i>	sim, ambas	concreto, factual, simples
96	<i>Interests special: theoretical</i>	133	<i>hypothetical</i>	hipotético, baseado em especulação	Thesaurus TT software, WordReference.com	<i>factual</i>	sim	factual
96	<i>Interests special: theoretical</i>	134	<i>conceptual</i>	conceitual	Thesaurus TT software, WordReference.com	<i>real, realistic</i>	sim	real
96	<i>Interests special: theoretical</i>	135	<i>speculative</i>	especulativo	Thesaurus TT software, WordReference.com	<i>proven, safe</i>	sim	não especulativo, provado
96	<i>Interests special: theoretical</i>	136	<i>idealistic</i>	idealista	Thesaurus TT software, WordReference.com	<i>realistic</i>	sim	realista
96	<i>Interests special: theoretical</i>	137	<i>imaginary</i>	hipotético, imaginário	Thesaurus TT software, WordReference.com	<i>factual, realistic</i>	sim	factual, realista
96	<i>Interests special: theoretical</i>	138	<i>academic</i>	acadêmico, teórico, ideal, intelectual	Thesaurus TT software, WordReference.com	<i>simple, plain, realistic, untaught</i>	sim	simples, realista, não-letrado
96	<i>Interests special: theoretical</i>	139	<i>conjectural</i>	conjectural	Thesaurus TT software, WordReference.com	<i>factual</i>	sim	factual
96	<i>Interests special: theoretical</i>	140	<i>ideal</i>	ideal, teórico	Thesaurus TT software, WordReference.com	<i>factual, realistic</i>	sim	factual, realista
97	<i>Interests wide</i>	141	<i>interested</i>	interessado	Thesaurus TT software	<i>interests narrow (Catell), uninterested</i>	sim, a segunda	desinteressado
98	<i>introspective</i>	142	<i>introspective</i>	introspectivo, tímido	-	<i>not introspective</i>	não	não-introspectivo
99	<i>intuitive</i>	143	<i>intuitive</i>	intuitivo	-	<i>logical (Catell), calculated, premeditated</i>	sim, as duas últimas	calculista, que medita as coisas antes da ação
100	<i>irritable</i>	144	<i>irritable</i>	irritável	-	<i>good tempered (Catell), pleasant</i>	sim, a últimas	agradável
101	<i>jealous</i>	145	<i>jealous</i>	ciumento, possessivo	-	<i>trusting</i>	sim	que confia no outro, seguro de si
102	<i>kind (by disposition)</i>	146	<i>kind-hearted</i>	bom, de boa natureza, humanitário, beneficente	Thesaurus TT software	<i>ruthless</i>	sim	sem piedade, cruel, brutal, inclemente

103	<i>kind (on principle)</i>	147	<i>kind</i>	amável, gentil	Thesaurus TT software	<i>unkind</i>	sim	indelicado
104	<i>laughterful</i>	148	<i>laughing</i>	risonho, sorridente	Thesaurus TT software	<i>mirthless (Catell) not laughing</i>	não	deprimido, melancólico, não-sorridente
105	<i>leading(not domineering)</i>	149	<i>leader</i>	líder	Thesaurus TT software	<i>follower, servant</i>	sim, ambas	seguidor, servente
105	<i>leading(not domineering)</i>	150	<i>influential</i>	influyente	Thesaurus TT software	<i>insignificant</i>	sim	insignificante
105	<i>leading(not domineering)</i>	151	<i>influencing</i>	quem influencia	Thesaurus TT software	<i>insignificant</i>	sim	insignificante
105	<i>leading(not domineering)</i>	152	<i>leading</i>	melhor que os demais, com mais excelência	-	<i>not leading</i>	não	sem excelência no que faz
105	<i>leading(not domineering)</i>	153	<i>governing</i>	dominante, primeiro, predominante	Thesaurus TT software	<i>inferior, subordinate</i>	sim, ambas	inferior, subordinado
106	<i>memory good</i>	154	<i>mindful</i>	ciente, consciente, diligente, que sabe todas as respostas	Webster Online, Merriam-Webster Online, WordNet 2.1 Browser e TT software Thesaurus, acessos de 17-01-12 a 20-01-2012	<i>unmindful</i>	sim	desatencioso, irrefletido
107	<i>mischievous</i>	155	<i>mischievous</i>	travesso, daninho, malicioso	-	<i>well behaved, obedient</i>	sim	bem comportado, obediente
108	<i>mulish</i>	156	<i>mulish</i>	obstinado	-	<i>not obstinate</i>	não	não-obstinado
109	<i>mystical</i>	157	<i>mystical</i>	místico (esotérico ou divino ou sombrio ou misterioso ou oculto)	-	<i>not mystical</i>	não	não-místico
110	<i>opinionated</i>	159	<i>opinionated</i>	teimoso, obstinado, cheio de opinião	-	<i>tolerant*</i>	sim	tolerante
111	<i>optimistic</i>	160	<i>optimistic</i>	otimista	-	<i>pessimistic</i>	sim	pessimista
112	<i>original</i>	161	<i>original</i>	original, autêntico	-	<i>not authentic</i>	não	não-autêntico
113	<i>painstaking</i>	162	<i>painstaking</i>	detalhista. Minucioso	-	<i>slipshod</i>	sim	deleixado, descuidado
114	<i>patient</i>	163	<i>patient</i>	paciente	-	<i>impatient</i>	sim	impaciente
115	<i>pedantic</i>	164	<i>pedantic</i>	preocupado demais com detalhes, pedante	-	<i>disordely (Catell), simple, imprecise</i>	sim	simple, impreciso

116	<i>persevering</i>	165	<i>persevering</i>	perseverante	-	<i>unsteadfast</i>	não	alterável, inconstante, não-firme, não-perseverante
117	<i>phantasying</i>	166	<i>fancy</i>	fantasioso	Webster Online e dicionário Webster	<i>plain</i>	sim	não-fantasiioso, claro
118	<i>physically active</i>	167	<i>athletic</i>	atletico	Thesaurus TT software	<i>inactive, unathletic</i>	sim a primeira	inativo
119	<i>pious</i>	168	<i>pious</i>	pio, moralizador	-	<i>impious</i>	sim	não-moralizador
120	<i>plaintive</i>	169	<i>plaintive</i>	lamentoso	-	<i>not plaintive</i>	não	não-lamentoso
121	<i>planful</i>	170	<i>prepared (well-planned)</i>	cheio de planos, preparado	Webster Online, Merriam-Webster Online, WordNet 2.1 Browser e TT software Thesaurus, acessos ede 17-01-12 a 20-01-2012	<i>unprepared</i>	sim	não-preparado
122	<i>poised</i>	171	<i>poised</i>	equilibrado	-	<i>awkward</i>	sim	desequilibrado, desajeitado
123	<i>polished</i>	172	<i>polished</i>	educado	-	<i>rough</i>	sim	rude, grosseiro
124	<i>practical</i>	173	<i>practical</i>	prático	-	<i>unrealistic</i>	sim	não-prático/realista
125	<i>proactive</i>	174	<i>active</i>	ativo	Thesaurus TT software	<i>inoperative</i>	sim	inoperante, sem proatividade
126	<i>prominent</i>	175	<i>prominent</i>	proeminente	-	<i>unimportant</i>	sim	insignificante
127	<i>pugnacious</i>	176	<i>pugnacious</i>	combativo, desafiador	-	<i>submissive</i>	sim	submisso
128	<i>reliable</i>	177	<i>reliable</i>	confiável	-	<i>undependable</i>	sim	não-confiável
129	<i>reserved</i>	178	<i>reserved</i>	reservado	-	<i>intrusive</i>	sim	intrusivo
130	<i>resilient</i>	179	<i>resilient</i>	resistente, que se mantém firme, que dá a volta por cima (rebounding)	-	<i>depressible (Ctell), delicate, not rebounding</i>	sim, as duas últimas	delicado, depressivo, que não dá a volta por cima
131	<i>responsive</i>	180	<i>responsive</i>	receptivo, responsivo (responde rápido)	-	<i>aloof (Catell), unresponsive</i>	sim, a segunda	indiferente, não-responsivo
132	<i>reverent</i>	181	<i>reverent</i>	reverente, reverencioso, respeitoso	-	<i>rebellious</i>	sim	rebelde, insubordinado

133	<i>sadistic</i>	182	<i>inhuman, inhumane</i>	inumano, sem compaixão, se satisfaz ouvindo os outros	Webster Online, Merriam-Webster Online, WordNet 2.1 Browser e TT software Thesaurus, acessos de 17-01-12 a 20-01-2012	<i>masochistic (Catell), self-punishing</i>	não	masoquista
134	<i>sarcastic</i>	183	<i>sarcastic</i>	sarcástico	-	<i>mild*</i>	sim	moderado, brando
135	<i>self-confident</i>	184	<i>self-confident</i>	auto confiante	-	<i>self-distrusting (Catell), insecure</i>	sim, a última	inseguro
136	<i>self-controlled</i>	185	<i>controlled</i>	controlado, auto-controlado	Thesaurus TT software	<i>uncontrolled</i>	sim	descontrolado
137	<i>selfish</i>	186	<i>selfish</i>	só pensa em si mesmo, egoísta	-	<i>self-denying*</i>	não	que nega a si mesmo, não pensa em si
138	<i>self-pitying</i>	187	<i>pitying</i>	suportivo, condolente	Dicionário Webster físico, Thesaurus TT software e thesaurus.com	<i>disregarding*</i>	não	que não se preocupa com nada, insensível
139	<i>self-respecting</i>	188	<i>elevated</i>	elevado, dignificado	Dicionário Webster físico e Thesaurus TT software, Thesaurus.com, WordNet 3.0 e websters-online-dictionary.org	<i>undignified</i>	sim	não-dignificado
139	<i>self-respecting</i>	189	<i>self-conscious</i>	auto-consciente	Dicionário Webster físico e Thesaurus TT software, Thesaurus.com, WordNet 3.0 e websters-online-dictionary.org	<i>not self-conscious</i>	não	não-consciente de si mesmo
139	<i>self-respecting</i>	190	<i>dignified</i>	dignificado	Dicionário Webster físico e Thesaurus TT software, Thesaurus.com, WordNet 3.0 e websters-online-dictionary.org	<i>undignified</i>	sim	não-dignificado
140	<i>self-respecting</i>	191	<i>self-satisfied</i>	satisfeito consigo mesmo	Dicionário Webster físico e Thesaurus TT software, Thesaurus.com, WordNet 3.0	<i>undignified</i>	sim	não-dignificado
141	<i>sensitive</i>	192	<i>sensitive</i>	sensível	-	<i>tough</i>	sim	duro, durão

142	<i>sentimental</i>	193	<i>sentimental</i>	sentimental	-	<i>hard-headed*</i>	sim	impassível, não-sentimental
143	<i>serious</i>	194	<i>serious</i>	sério	-	<i>frivolous</i>	sim	frívolo, superficial, bobo, não-sério
144	<i>shrewd</i>	195	<i>shrewd</i>	astuto, sagaz, perspicaz	-	<i>naive</i>	sim	não-refinado/sagaz, ingênuo
145	<i>slandorous</i>	196	<i>slandorous</i>	caluniador, difamador	-	<i>complimentary, harmless*</i>	não	elogioso, manerado
146	<i>sleeps well</i>	197	<i>fresh</i>	revigorado	Webster Online e dicionário Webster	<i>exhausted</i>	sim	esgotado (mentalmente, emocionalmente)
147	<i>slow (temperamentally)</i>	198	<i>slow</i>	lento, moderado para reagir	Thesaurus TT software, thesaurus.com e wordreference.com	<i>quick</i>	sim	rápido, sagaz, vivo
148	<i>sociable I (meeting people)</i>	199	<i>sociable</i>	sociável, social	Dicionário Webster físico e Thesaurus TT software	<i>shy</i>	sim	tímido
149	<i>sociable II (regarious)</i>	200	<i>gregarious</i>	gregário, se junta a pessoas, agrega pessoas	Dicionário Webster físico e Thesaurus TT software	<i>unfriendly</i>	sim	não-amigável
150	<i>sophisticated</i>	201	<i>sophisticated</i>	sofisticado/culto	Thesaurus TT software	<i>unsophisticated</i>	sim	não-sofisticado/culto
151	<i>sour</i>	202	<i>sour</i>	azedo, desagradável	-	<i>sweet</i>	sim	doce, agradável
152	<i>stable emotionally</i>	203	<i>stable</i>	estável	Dicionário Webster físico e Thesaurus TT software	<i>unstable</i>	sim	instável
153	<i>strong in personality</i>	204	<i>strong</i>	bravo, corajoso, destemido, valente, valoroso, intrépido, firme, convincente, influente	Dicionário Webster físico e Thesaurus TT software	<i>weak</i>	sim	não-corajoso, sem valentia, não-valoroso, sem firmeza, sem influência
154	<i>subjective</i>	205	<i>subjective</i>	subjetivo	Thesaurus TT software	<i>guided by reality (Catell), realistic</i>	sim a segunda	guiado pela realidade, realista
155	<i>suggestible</i>	206	<i>advisable</i>	aconselhável	-	<i>inadvisable</i>	sim	desaconselhável, inaconselhável
156	<i>thoughtful</i>	207	<i>thoughtful</i>	pensativo	Thesaurus TT software	<i>unreflective*</i>	não	não-reflexivo
157	<i>timid (disposition)</i>	208	<i>timid</i>	tímido	Thesaurus TT software	<i>adventurous*</i>	sim	aventureiro

158	<i>treacherous</i>	209	<i>perfidious</i>	desleal, pérfido, inconfiável, falso	thefreedictionary.com, Thesaurus TT software,	<i>untreacherous, loyal</i>	sim, a última	leal
159	<i>versatile</i>	210	<i>many-sided</i>	com muitos lados, múltiplo, multifacetado, adaptável, eclético	thefreedictionary.com, Thesaurus TT software, WordNet 3.0	<i>narrow, limited, restricted</i>	sim, ambas	limitado, restrito
159	<i>versatile</i>	211	<i>multifarious</i>	com muitos lados, múltiplo, multifacetado, adaptável, eclético	thefreedictionary.com, Thesaurus TT software, WordNet 3.0	<i>narrow, limited, restricted</i>	sim, ambas	limitado, restrito
159	<i>versatile</i>	212	<i>multiple</i>	com muitos lados, múltiplo, multifacetado, adaptável, eclético	thefreedictionary.com, Thesaurus TT software, WordNet 3.0	<i>narrow, limited, restricted</i>	sim, ambas	limitado, restrito
159	<i>versatile</i>	213	<i>eclectic</i>	eclético	thefreedictionary.com, Thesaurus TT software, WordNet 3.0	<i>narrow, limited, restricted</i>	sim, ambas	limitado, restrito
160	<i>wise</i>	214	<i>wise</i>	sábio	-	<i>foolish*</i>	sim	bobó, tolo
161	<i>worrying</i>	215	<i>worried</i>	preocupado	Thesaurus TT software	<i>unworried, undisturbed</i>	sim, a segunda	não-preocupado
162	<i>vivacious</i>	216	<i>vivacious</i>	-	-	-	-	-
163	<i>witty</i>	217	<i>witty</i>	-	-	-	-	-
164	<i>tactful</i>	218	<i>tactful</i>	-	-	-	-	-
165	<i>talkative</i>	219	<i>talkative</i>	-	-	<i>taciturn</i>	-	-
166	<i>temperate</i>	220	<i>temperate</i>	-	-	<i>extreme (in schizothyme sense)</i>	-	-
167	<i>vindictive</i>	221	<i>vindictive</i>	-	-	<i>unrepentful</i>	-	-
168	<i>wandering</i>	222	<i>wandering</i>	-	-	<i>settling down</i>	-	-
169	<i>self-deceiving (Não incluída)</i>	-	<i>misguided</i>	enganado	Dicionário Webster físico e Thesaurus TT software	<i>right</i>	sim	certo
170	<i>loyal (Não incluída)</i>	-	-	-	-	-	-	-
171	<i>Mature (emotional)- (Não incluída)</i>	-	-	-	-	-	-	-

6.1.1 Critérios para substituição de algumas das 171 palavras de Cattell, por palavras similares no *thesaurus* utilizado.

Das 171 palavras a que chegou CATTELL (1943b), algumas não existiam no *thesaurus* utilizado neste trabalho. Para superar este problema recorreu-se à substituição

destas palavras ausentes no banco utilizado por outras palavras próximas em sentido aos traços, originalmente propostos por este autor.

Em alguns casos, mais de uma palavra candidata a substituta de alguma outra palavra foi considerada na lista final a ser utilizada (para isso consultou-se outros dicionários - físicos, no caso do dicionário Webster - e eletrônicos - na *internet* e o *thesaurus* wordnet instalado como *software* - para se verificar que outras palavras estavam associadas a estas palavras faltantes e se selecionou a que mais se aproximava em sentido destas palavras faltantes). Assim, considerando-se as palavras originais existentes no *thesaurus* mais as substitutas (próximas em sentido) das palavras ausentes) chegou-se a um conjunto final de 222 adjetivos-traços.

A tabela acima apresenta as palavras inicialmente propostas por Catell, os termos que efetivamente foram utilizados nas aplicações de clusterização a seguir, seus antônimos e traduções.

No caso de *musical*, ela substituiu duas palavras originais de Catell: *Abilities: musical aptitude* e *Interests special: music*.

Neste sentido, verifica-se que a palavra *hypochondriac* foi substituída por três palavras, cada uma com um significado possível desta palavra: *anxious*, *neurotic* e *splenetic*. Assim, na busca de palavras próximas a *hypochondriac*, entendeu-se que esta palavra podia denotar tanto alguém ansioso (*anxious*) como alguém neurótico (*neurotic*). Por esta razão se adotou as três palavras como substitutas desta palavra inicial.

No caso da palavra *self-respecting* (ausente no TTT), optou-se por substituí-la por: *self-conscious*, pois, o autorrespeito é um caso particular de autoconsciência. Outras candidatas a substituta dessa palavra também foram incluídas: *elevated*, *dignified* e *self-satisfied*.

No caso do termo *strong in personality*, escolheu-se a palavra *strong* como substituta. Isso se deveu ao fato de que a palavra *strong* é mais ampla do que a palavra inicialmente proposta.

De modo similar, *Emotional II* foi substituída por *convivial* e *affable*.

A palavra *loyal*, proposta inicialmente por Catell, não foi incluída na lista das palavras utilizadas nas aplicações com as 222 palavras, pois, ela é semelhante semanticamente a *devout*.

A palavra *neurotic* (sugerida por Cattell) foi utilizada. Contudo, como visto anteriormente, *hypochondriac* possui como um dos seus sentidos justamente o de ser neurótico. Assim, *neurotic* também é uma palavra substituta de *hypochondriac*.

A palavra *mature*, também originalmente proposta por Cattell, não foi incluída no conjunto das 222 palavras, pois, de acordo com o *thesaurus* da TT *software* ela está relacionada com as palavras *sophisticated*, já selecionada anteriormente.

No caso da palavra *self-deceiving* (que se auto engana), não se encontrou palavras substitutas que se aproximassem suficientemente do sentido original dessa palavra.

Por restrições de tempo não se pesquisou exaustivamente palavras que provavelmente existam no *thesaurus* utilizado cumprindo significados mais próximos ao deste termo.

No caso do termo *self-pitying*, o mesmo foi substituído por *pitying* (suportivo, condolente), que não corresponde exatamente aos sentidos originais do traço proposto por Cattell.

No caso da palavra *self-respecting*, também proposta por Cattell, como a mesma não está presente na versão do *thesaurus* utilizada, se resolveu adotar outras palavras com significados bem correlacionados ao significado desta expressão. Neste sentido foram consideradas como substitutas as palavras *elevated* (elevado, dignificado), *dignified* (dignificado) e *self-conscious* (autoconsciente ou com consciência de si).

Esta última expressão (*self-conscious*) foi considerada, pois se pode entender que a autoconsciência é um aspecto comportamental mais amplo que o de se auto respeitar (que pode ser entendido como uma consequência da autoconsciência).

De acordo com o sítio da *internet* thefreedictionary.com, a palavra *self-conscious* significa “Aware of oneself as an individual or of one's own being, actions, or thoughts”, podendo incluir também como um caso particular dessa característica a consciência de autorrespeito às próprias características do indivíduo que seja autoconsciente.

Por este motivo, caso se fosse escolher uma única palavra (ou expressão) para substituir *self-respecting*, ela seria *self-conscious*, por se tratar de um traço mais exaustivo, no sentido de retratar ou representar aspectos comportamentais mais amplos. No entanto, como boa parte dessas palavras candidatas a substitutas de *self-respecting* se encontravam relacionadas (ligadas) entre si no *thesaurus* utilizado resolveu-se adotar todas elas para fins de teste da capacidade das diferentes metodologia de clusterização de agrupar palavras fortemente relacionadas em *clusters* comuns.

Como se verá a seguir, este mesmo procedimento (incluir no conjunto final de palavras para teste das metodologias de clusterização, todas as palavras – ou expressões-substitutas das palavras ausentes) foi adotado mais de uma vez na construção do conjunto de palavras utilizado como teste das citadas metodologias; sempre por esse mesmo motivo.

Vale esclarecer aqui que, em muitas aplicações com traços de personalidade provavelmente uma única palavra substituta (a mais exaustiva em termos de cobertura de aspectos comportamentais diferentes) seria escolhida para cada palavra ausente no banco de dados adotado. No entanto, por se tratar de uma ilustração das metodologias nesse *thesaurus* aberto, adotou-se este procedimento.

No caso da palavra *Interests special: artistic*, ausente no banco de dados utilizado, ela foi substituída por outra que representa um aspecto, sentido (potencial) particular da mesma: *decorative*. Pode-se assim se depreender que pessoas que tenham interesses artísticos sejam também (ao menos algumas delas) decorativas e estilosas (ou preocupadas com questões relacionadas à decoração e estilo).

No caso do traço *Abilities: mechanical aptitude* (aptidão mecânica), não se encontrou esta palavra no TTT empregado, mas foram encontradas diversas palavras com significados relacionados ao deste traço. Em função disso resolveu-se substituir esta palavra por diversas outras, representando aspectos da personalidade correlacionados com o desta palavra. A primeira palavra substituta escolhida foi *mechanical*. De acordo com o *thesaurus online thesaurus.com*, *mechanical* pode significar também alguém sem emoção, que não demonstra emoção alguma (*emotionless, cold*), assim como alguém que age de forma habitual, instintiva e involuntária (*involuntary*), alguém que age somente de forma programada (*programmed*) e padronizada (*standarized*).

Mais especificamente foram consideradas as seguintes palavras como substitutas desta: *programmed* (programado, automatizado, quem faz as coisas automaticamente, sem pensar), *standardized* (padronizado, sem criação, alguém que faz coisas da mesma forma, segundo normas), *perfunctory* (superficial, perfunctório, desinteressado, sem entusiasmo e sem pensar) e *instinctive* (instintivo).

Novamente, embora os estudos psicoléxicos na área de traços de personalidade tenham mostrado de que a escolha dos traços a serem utilizados em uma dada aplicação deva ser guiada também pela abrangência dos aspectos comportamentais cobertos por elas, resolveu-se incluir também, nas aplicações a seguir, as duas últimas palavras do parágrafo anterior (*programmed* e *standarized*). A razão para isto é a mesma da alegada anteriormente:

incluiram-se estas últimas seria testar a proficiência com que os diferentes métodos de clusterização são capazes de organizar-nos mesmos clusters palavras similar em sentido.

Pelo mesmo motivo, se resolveu incorporar também no conjunto final para teste das metodologias, as palavras *speaking* e *unwritten* (relacionadas com *vocal*).

Também pela mesma razão, o traço *versatile* foi substituído pelas palavras *many-sided*, *multifarious eclectic* e *multiple*. Idem com relação às palavras *deft*, *skilled*, *skilful*, *dexterous* e *proficient* (que foram todas consideradas substitutas do traço *abilities: dextery*, ausente do *thesaurus*).

O mesmo acontece com as palavras *rational* e *logical* (substituindo o traço *abilities: reasoning*). E com relação às palavras *lexical*, *verbal*, *vocal*, *speaking*, *unwritten* e *open-hearted*, que entraram todas no conjunto final de palavras para as aplicações com os métodos de clusterização; estas palavras substituíram o traço *abilities: verbal attitude* (ausente do TTT, por ser um termo construído, composto).

O mesmo vale para *discriminating* e *decorative* (substituindo *Interests special: artistic*) e *rational* e *logical* (substituindo o traço *abilities: reasoning*). Idem para as palavras *economical*, *frugal*, *provident*, *thrifty* e *sparing*, substituindo o traço *interests special: economic*.

Idem para as palavras *leading*, *influential*, *influencing* e *leader*, substitutas de *Leading (not Domineering)*.

Idem para as palavras *abstract*, *hypothetical*, *conceptual*, *speculative*, *idealistic*, *imaginary*, *academic*, *conjectural* e *ideal*, que entraram todas substituindo o traço *Interests special: theoretical*.

Com relação ao traço *abilities: spatial, visual ability*, não se encontrou uma palavra substituta que traduzisse bem seu significado. No entanto encontraram-se no TTT palavras que, embora semelhantes em sua construção léxica, possuem significados diferentes das palavras inicialmente propostas, mas importantes em termos dos aspectos comportamentais a que se referem. São elas: *visionary*, *visible*, *visual*, *observing* e *observer*.

O traço *agoraphobic* (agorafóbico, ou quem sofre de agorafobia, ou seja, fobia de lugares abertos), inicialmente proposto, não se encontrava no *thesaurus* utilizado e foram substituídos pelas palavras *frightened*, *panicky* e *coward*. Incluiu-se também a palavra *reclusive*, pois pessoas que sejam agorafóbicas tendem também a serem pessoas reclusas. Isto também foi verificado por meio das associações entre palavras envolvendo o termo *agoraphobic*.

Com relação ao termo *claustrophobic* (claustrofóbico), o mesmo foi substituído pela palavra *afraid*, pois também não se encontrava no TTT e não se encontrou nenhum outro termo que melhor aproximasse seu significado.

Com relação ao traço *Interests Special: economic*, obviamente não se encontrou o mesmo no TTT por se tratar de uma expressão composta de vários termos. Na busca de substitutos deste traço selecionou como candidatos as seguintes palavras: *economical* (econômico), *frugal* (frugal), *provident* (providente), *thrifty* (sóbrio, frugal), *sparing* (modesto e frugal). Novamente, como estas palavras estavam diretamente relacionadas entre si no *thesaurus* adotado, resolveu-se considerar todas estas nas aplicações a seguir com propósitos de se testar as metodologias de clusterização nas suas capacidades de agrupar palavras proximamente relacionadas nos mesmo *clusters*.

Novamente pelos mesmos motivos consideraram-se as palavras *social*, *informative*, *hospitable*, *amiable* e *cordial* todas como substitutas do traço *Interests special: social*.

Com relação ao traço *Abilities: physical endurance*, apresentado por Catell, como se trata de um nome (e não de um adjetivo) não se conseguiu achar um adjetivo-traço que o pudesse substituir. Esta palavra foi substituída pela expressão *in shape* (em forma), presente no TTT, que também é um nome, e não um adjetivo.

O traço *loyal*, proposto por Catell na sua lista de 171 palavras não foi incluído no conjunto final das 222 palavras, apesar de existir no *thesaurus* eletrônico utilizado. A razão para a sua não inclusão reside basicamente no fato de que se omitiu esta palavra de forma não intencional no início das aplicações computacionais e só se percebeu a sua ausência quando todos os resultados finais com todas as metodologias já estavam analisados. Contudo, novamente como se tratam de aplicações ilustrativas das metodologias adotadas, sua ausência não implica em perdas para se atingir os objetivos deste trabalho.

Apesar desta exclusão, *loyal* está presente como antônimo da palavra *perfidious* (que por sua vez substituiu a palavra *treacherous*).

O mesmo ocorre com a palavra *mature* (*emotional*). Esta palavra não está presente no TTT, mas a palavra *mature* (que poderia substituí-la) existe no *thesaurus* utilizado. Contudo, esta última também não foi incluída no conjunto final de palavras selecionadas para as aplicações de clusterização, pelo mesmo de *loyal* não ter sido incluída (exclusão acidental ou não intencional). Novamente, sua não inclusão não impacta no comprometimento dos resultados obtidos e dos objetivos traçados para este trabalho.

Vale a pena ressaltar também que, na tabela apresentada acima, em alguns casos de antônimos das palavras propostas por Cattell, não foram encontradas substitutas presentes no TTT, verdadeiramente próximas em sentido às propostas por este autor. Nestes casos, as palavras antônimas originais do autor foram mantidas no quadro, marcadas com um asterisco para realçar suas ausências no thesaurus adotado neste trabalho.

Em alguns poucos casos, algumas palavras candidatas a substitutas de palavras originalmente propostas por Cattell eram tão próximas em sentido dessas palavras originalmente propostas que se decidiu adotar as duas simultaneamente, mesmo sabendo-se que significam praticamente a mesma coisa.

A razão para se proceder a estas substituições reside primeiramente no fato de que, apesar de algumas palavras substitutas não significarem exatamente a mesma coisa das palavras originalmente propostas por Cattell, precisava-se de palavras que existissem no thesaurus utilizado para que estas palavras possam estar ligadas entre si e com outras e desta forma se pudessem usar os métodos de clusterização aqui utilizados.

Uma alternativa à substituição das palavras inexistentes no banco de dados seria, com o auxílio de um linguista qualificado, alterar o *thesaurus* original, anexando estas palavras ao TTT, elegendo-se as melhores palavras com as quais as mesmas deveriam estar ligadas.

Desta forma se estaria incluindo palavras no TTT e se incluindo também vértices e ligações no grafo correspondente às associações entre estas palavras. Contudo como não se dispunha de tempo e de formas de contratação deste profissional, além do objetivo aqui ser somente ilustrar as possibilidades com os desenvolvimentos desta tese, se optou pela substituição de palavras próximas em sentido as palavras faltantes no banco).

Na fonte da tabela relativa as palavras substituídas, as originais, as traduções e os antônimos colocar que para as substituições das palavras originais foram consultadas associações de palavras nos dicionários Webster online, Merriam-Webster Online, Wordnet 2.1 *Browser software* e *TT Software Thesaurus* no período de 17-01-12 a 20-01-12.

Com estas escolhas, procedeu-se à limpeza da planilha Excel contendo o TTT todo, de forma que somente estas 222 palavras permanecessem (assim como as associações entre elas). A partir desta nova planilha, geraram-se todos os relatórios necessários para entrada nos diferentes algoritmos utilizados (tais como listas de adjacência do grafo associado às ligações entre as palavras selecionadas, assim como sua matriz de adjacência, a matriz de Markov para entrada no algoritmo MCL, e os vetores representando os vértices desse grafo).

6.1.2 Análise do grafo que representa as associações entre as 222 palavras selecionadas

Com as palavras assim escolhidas, o grafo que representa estas palavras e as associações entre elas, é formado por 54 componentes conexos.

Dessas componentes conexas, tem-se uma grande componente formado por 145 vértices. A segunda maior (componente 6) tem 11 vértices. A seguir tem-se a componente 10 formada por 4 vértices. As demais componentes com cardinalidade superior a um são formadas por 3 ou 2 vértices.

As demais componentes (43 no total, o correspondente a 19,37% do total de vértices deste grafo) são formadas por vértices isolados, sem ligações com os demais do banco de dados assim delimitado (isto é, considerando-se somente estas 222 palavras e as ligações entre elas).

Portanto este grafo não é uma árvore. Trata-se de um grafo desconexo e com ciclos. Para efeitos de utilização dos dois algoritmos de clusterização empregados nesta tese, este grafo foi considerado não direcionado.

A seguir apresenta-se uma tabela listando estas 222 palavras por componente conexo (grupo) a que pertencem.

Tabela 23: As 222 palavras consideradas e seus respectivos componentes conexos.

COMPONENTE CONEXO	Palavra
1	<i>frank</i>
1	<i>open-hearted</i>
1	<i>devout</i>
1	<i>honest</i>
1	<i>religious</i>
1	<i>pious</i>
1	<i>reverent</i>
1	<i>clever</i>
1	<i>dexterous</i>
1	<i>witty</i>
1	<i>skilful</i>
1	<i>deft</i>
1	<i>skilled</i>
1	<i>prepared</i>

1	<i>proficient</i>
1	<i>professional</i>
1	<i>alert</i>
1	<i>fresh</i>
1	<i>original</i>
1	<i>economical</i>
1	<i>frugal</i>
1	<i>provident</i>
1	<i>thrifty</i>
1	<i>sparing</i>
1	<i>discriminating</i>
1	<i>imaginative</i>
1	<i>aesthetic</i>
1	<i>polished</i>
1	<i>informative</i>
1	<i>hospitable</i>
1	<i>poised</i>
1	<i>social</i>
1	<i>sociable</i>
1	<i>gregarious</i>
1	<i>convivial</i>
1	<i>genial</i>
1	<i>cordial</i>
1	<i>friendly</i>
1	<i>amiable</i>
1	<i>affable</i>
1	<i>political</i>
1	<i>cooperative</i>
1	<i>coherent</i>
1	<i>cosmopolitan</i>
1	<i>impartial</i>

1	<i>cheerful</i>
1	<i>resilient</i>
1	<i>optimistic</i>
1	<i>affectionate</i>
1	<i>kind</i>
1	<i>responsive</i>
1	<i>kind-hearted</i>
1	<i>charming</i>
1	<i>grateful</i>
1	<i>hearty</i>
1	<i>generous</i>
1	<i>rational</i>
1	<i>analytical</i>
1	<i>advisable</i>
1	<i>intelligent</i>
1	<i>logical</i>
1	<i>wise</i>
1	<i>shrewd</i>
1	<i>hard</i>
1	<i>practical</i>
1	<i>sophisticated</i>
1	<i>constructive</i>
1	<i>eclectic</i>
1	<i>multifarious</i>
1	<i>strong</i>
1	<i>alcoholic</i>
1	<i>energetic</i>
1	<i>courageous</i>
1	<i>influential</i>
1	<i>in shape</i>
1	<i>active</i>

1	<i>inflexible</i>
1	<i>severe</i>
1	<i>exact</i>
1	<i>ambitious</i>
1	<i>ascetic</i>
1	<i>serious</i>
1	<i>enterprising</i>
1	<i>industrious</i>
1	<i>austere</i>
1	<i>formal</i>
1	<i>many-sided</i>
1	<i>multiple</i>
1	<i>thoughtful</i>
1	<i>mindful</i>
1	<i>cautious</i>
1	<i>visionary</i>
1	<i>ideal</i>
1	<i>hypothetical</i>
1	<i>reclusive</i>
1	<i>reserved</i>
1	<i>inhibited</i>
1	<i>dignified</i>
1	<i>elevated</i>
1	<i>eloquent</i>
1	<i>conventional</i>
1	<i>mechanical</i>
1	<i>programmed</i>
1	<i>observing</i>
1	<i>doubtful</i>
1	<i>self-conscious</i>
1	<i>self-satisfied</i>

1	<i>self-confident</i>
1	<i>standardized</i>
1	<i>instinctive</i>
1	<i>intuitive</i>
1	<i>speculative</i>
1	<i>conjectural</i>
1	<i>academic</i>
1	<i>boastful</i>
1	<i>conceited</i>
1	<i>arrogant</i>
1	<i>egotistical</i>
1	<i>imitative</i>
1	<i>autocratic</i>
1	<i>opinionated</i>
1	<i>selfish</i>
1	<i>impulsive</i>
1	<i>emotional</i>
1	<i>idealistic</i>
1	<i>imaginary</i>
1	<i>subjective</i>
1	<i>abstract</i>
1	<i>affected</i>
1	<i>vivacious</i>
1	<i>sentimental</i>
1	<i>conceptual</i>
1	<i>worried</i>
1	<i>frightened</i>
1	<i>irritable</i>
1	<i>enthusiastic</i>
1	<i>excitable</i>
1	<i>panicky</i>

1	<i>timid</i>
1	<i>sour</i>
1	<i>sensitive</i>
1	<i>high-strung</i>
1	<i>anxious</i>
1	<i>afraid</i>
1	<i>coward</i>
2	<i>fancy</i>
3	<i>fastidious</i>
4	<i>flattering</i>
5	<i>preventive</i>
6	<i>animal</i>
6	<i>prominent</i>
6	<i>bodily</i>
6	<i>physical</i>
6	<i>visible</i>
6	<i>curious</i>
6	<i>leading</i>
6	<i>corporal</i>
6	<i>interested</i>
6	<i>corporeal</i>
6	<i>perceivable</i>
7	<i>acquisitive</i>
8	<i>amorous</i>
9	<i>argumentative</i>
9	<i>pugnacious</i>
10	<i>assertive</i>
10	<i>vocal</i>
10	<i>speaking</i>
10	<i>unwritten</i>
11	<i>athletic</i>

12	<i>closed</i>
13	<i>complaining</i>
13	<i>cynical</i>
13	<i>sarcastic</i>
14	<i>conscientious</i>
14	<i>painstaking</i>
14	<i>tactful</i>
15	<i>contented</i>
16	<i>controlled</i>
16	<i>patient</i>
16	<i>persevering</i>
17	<i>decorative</i>
17	<i>musical</i>
18	<i>easygoing</i>
19	<i>eccentric</i>
20	<i>effeminate</i>
21	<i>evasive</i>
22	<i>gluttonous</i>
23	<i>governing</i>
24	<i>habituated</i>
25	<i>headstrong</i>
25	<i>mulish</i>
26	<i>hurried</i>
27	<i>independent</i>
28	<i>influencing</i>
29	<i>inhuman</i>
30	<i>introspective</i>
31	<i>jealous</i>
32	<i>laughing</i>
33	<i>leader</i>
34	<i>lexical</i>

34	<i>verbal</i>
35	<i>mischievous</i>
36	<i>mystical</i>
37	<i>neurotic</i>
38	<i>observer</i>
39	<i>pedantic</i>
40	<i>perfidious</i>
41	<i>perfunctory</i>
42	<i>pictorial</i>
43	<i>pitying</i>
44	<i>plaintive</i>
45	<i>reliable</i>
45	<i>stable</i>
46	<i>slanderous</i>
47	<i>slow</i>
48	<i>splenetic</i>
49	<i>sportive</i>
50	<i>talkative</i>
51	<i>temperate</i>
52	<i>vindictive</i>
53	<i>visual</i>
54	<i>wandering</i>

A seguir se apresenta o número de vértices (palavras) por componente conexo do grafo).

Tabela 24: Número de palavras por componente conexo do grafo

COMPONENTE CONEXO	Número de Traços por Componente
1	145
2	1
3	1
4	1

5	1
6	11
7	1
8	1
9	2
10	4
11	1
12	1
13	3
14	3
15	1
16	3
17	2
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	2
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	2
35	1

36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	2
46	1
47	1
48	1
49	1
50	1
51	1
52	1
53	1
54	1
TOTAL	222

Analisando-se a tabela acima, constata-se que o grupo 1 corresponde ao componente conexo com mais palavras (145 palavras) no grafo que representa as relações entre as palavras consideradas.

O grupo 6 corresponde ao segundo maior componente conexo (11 palavras). O grupo 10 corresponde ao terceiro maior (com 4 traços). A seguir têm-se os grupos 13, 14 e 16 (todos com 3 traços).

Os grupos 9, 17, 25, 35 e 47 São formados todos por 2 palavras.

Todos os demais grupos são formados por palavras que ficaram isoladas com a escolha de somente essas 222 palavras, no banco de dados.

A representação gráfica desta rede de palavras, conjuntamente com a identificação e destaque visual dos *clusters* encontrados pelos algoritmos empregados são uma forma útil de representação destes TP, pois:

- Apresentam de forma clara como estes traços se relacionam, segundo a linguagem natural;
- Organiza e apresenta estes traços de acordo com suas proximidades (dentro de um sentido) semânticas e quanto ao comportamento a que se referem e;
- Provê modos de rápida identificação das regiões (*clusters*) de TP.

Os gráficos dos clusters encontrados com ambas as metodologias estão disponíveis no anexo do trabalho e foram gerados no *software* R versão 2.12.2 com o pacote Igraph, capaz de plotar gráficos de grafos.

6.2. DETALHES DE IMPLEMENTAÇÃO

Faz-se necessário iniciar esta sessão reforçando-se que uma boa solução de clusterização deve necessariamente levar em consideração as particularidades do problema abordado e características desejáveis para os *clusters* a serem definidos, que podem ser estabelecidas a priori para o problema.

Neste sentido, no que se refere ao problema de clusterização de grafos em questão (representando o problema de clusterização das 222 palavras escolhidas em um *thesaurus*) considerou-se uma boa solução de clusterização aquela nas quais não se tivesse uma parcela grande de palavras em *clusters* com cardinalidade unitária (isto é, formados só por uma única palavra) assim como não se observasse muitos *clusters* formados por um percentual muito elevado de palavras.

Assim determinou-se *a priori* que, descontando-se os vértices totalmente isolados do grafo, não mais do que 20% dos *clusters* encontrados deveriam ser formados por somente uma única palavra; e não se deveria aceitar *clusters* que tivessem 30 palavras - o corresponde a cerca de 13,51% dos dados a serem clusterizados - ou mais).

Estas regras adotadas visam garantir que os *clusters* obtidos pudessem ser úteis em um processo de organização (ao menos inicial) de palavras (mais especificamente de traços de personalidade), de modo que muitos *clusters* formados somente por uma única palavra seriam pouco úteis neste processo (pois se estaria perdendo a oportunidade de se agrupar palavras similares e compartilhando entre si sentidos mais amplos, desde que agrupadas nos mesmos *clusters*).

Do mesmo modo, *clusters* formados por muitas palavras tenderiam a reunir palavras já mais distantemente relacionadas entre si no *thesaurus*. Isto resultaria na possibilidade de que as ideias gerais presentes em tais *clusters* já fossem mais heterogêneas.

Fundamentalmente, todos os *clusters* devem conter vértices com caminho entre si, de modo a se respeitar esta condição fundamental para a definição de *clusters* em grafos.

Como visto anteriormente, o grafo em questão (que modela a rede de relacionamentos entre as palavras, no *thesaurus*) é desconexo, sendo composto, como já visto por 54 componentes conexos definidos a partir dos seus 222 vértices. Destes 54 componentes conexos, 43 são formados por vértices isolados (isto representa 19,37% do total de palavras utilizadas).

Uma componente (a componente 6) é formada por 11 elementos. O componente conexo 1, o maior de todos, é formado por 145 vértices. O componente 10 é formado por 4 elementos. O restante dos componentes conexos são formados por 2 ou 3 elementos.

Este grafo é cíclico, pois possui componentes conexos que apresentam ciclos. Considerou-se, nos dois métodos de clusterização utilizados, tal grafo com sendo não orientado.

Como neste grafo, diversos componentes conexos é composto por um único vértice isolado, este fato deve ser levado em consideração, pois os *clusters* definidos pelas duas metodologias aqui utilizadas não podem conter elementos desconexos (como já visto anteriormente). Isto obriga que soluções aceitáveis de clusterização devam formar *clusters* dedicados para estes vértices.

Adicionalmente, como já antecipado anteriormente, uma solução de clusterização adequada deve ser composta, para os componentes conexos formados por diversos vértices, por *clusters* (ou pelo menos para a maioria deles) formados por mais de um elemento; pois dessa forma esse estaria obtendo uma forma útil de organização dessas palavras, tirando proveito das propriedades de *clusters* de palavras em *thesaurus* (em especial a propriedade de que as palavras em tais *clusters* compartilham de sentidos/ideias em comum; presentes nesses *clusters*) já mencionadas anteriormente.

Por outro lado, esses *clusters* (definidos nos componentes conexos do grafo com maior quantidade de vértices) não devem conter um número substancialmente grande dos vértices do grafo.

Clusters formados por números substancialmente grandes de vértices do grafo reunirão vértices distantes uns dos outros no mesmo, o que implicaria em agruparem traços que se referem a aspectos mais afastados da personalidade humana. Seria mais adequado reunir traços mais similares em *clusters* comuns. Deste modo seria desejável haver certo grau de balanceamento do número de palavras nos *clusters* definidos (ou ao menos que a

solução provida pelos algoritmos utilizados forneça um balanceamento possível desse número, entre *clusters*, de acordo com a lógica de identificação desses agrupamentos utilizada por cada um desses algoritmos).

Adicionalmente, verificando-se que o grau de conectividade entre nós do grafo varia de região a região do mesmo (em particular no componente conexo 1, o maior de todos) deve-se aceitar que o tamanho dos *clusters* finais reflita o grau de conectividade da região do componente conexo do grafo a que se refere.

Adicionalmente, também seria desejável que ao menos em parte dos *clusters* encontrados se verificasse alta conectividade intra *cluster* e menor conectividade inter *clusters* (pelos motivos já apresentados anteriormente).

Assim, as escolhas das configurações dos dois algoritmos de clusterização empregados levaram em consideração estas verificações, que serão a seguir detalhadas.

6.2.1 Redes SOM

As redes neurais de Kohonen utilizadas rodaram no software MATLAB versão R2010b 64bits, conjuntamente com a *toolbox Neural Networks* desta versão.

- **Arquitetura do mapa topológico utilizado:** Escolheu-se um mapa topológico para a rede neural de arquitetura hexagonal e com arranjo não aleatório (mais precisamente esse arranjo é composto de 10 linhas de hexágonos e 20 colunas de hexágonos). Escolheu-se tal arranjo, pois o número de hexágonos não era muito diferente da quantidade de itens sendo clusterizados (222 palavras).

- **Medida de distância utilizada:** Adotou-se como medida de distância entre os elementos sendo clusterizados (ou mais particularmente das representações dos mesmos como vetores) e os modelos - os representando - apresentados pela rede, a distância Euclidiana, considerada a mais adequada para o problema e também se levando em consideração o pequeno porte do mesmo. Problemas de clusterização utilizando quantidades elevadas de vértices em um grafo poderão impor a utilização de outra medida de distância – já que a distância Euclidiana é mais cara computacionalmente.

- **Taxa de aprendizagem utilizada:** Adotou-se como taxa de aprendizagem final para a rede o valor de 0,6.

- **Número de épocas:** Neste trabalho se adotou a esquemática proposta por COTTRELL et all (2005) no que se refere ao modo de determinação do número de épocas (repetições) a ser utilizado no treinamento da rede neural de Kohonen. Segundo esta

esquemática, se adotaria como número de épocas o valor correspondente a cinco vezes o tamanho da amostra utilizada no treinamento da rede.

Como se utilizou um conjunto de 222 vetores no treinamento da mesma se definiu como o número de épocas o valor de 1000; assim se definiu 1000 épocas a serem apresentadas à rede (este número aproxima para um valor arredondado o valor de 1110 épocas, que seria o valor exato correspondente à operação 5×222).

Deve-se observar também que, realizaram-se diversos treinamentos da rede SOM com diversas combinações de parâmetros (incluindo diversas combinações de valores para taxas de aprendizagem, raio da vizinhança a partir de nós do mapa, o número de interações necessárias para que o raio destas vizinhanças se reduzisse a uma unidade e para o número de épocas); sempre se mantendo fixos os vetores codificados e a arquitetura do mapa topológico - mantida em 20×10 hexágonos -.

Com os repetidos treinamentos da rede, observou-se que a partir de 1000 épocas os resultados finais das clusterizações passavam a apresentar sensível menor variabilidade (em termos de mudanças de alocação das palavras nas regiões do mapa); de modo que este valor para o número de épocas foi mantido fixo nos treinamentos posteriores.

- **Valor inicial para o tamanho das vizinhanças:** se adotou como valor inicial para este parâmetro o valor de 15.

- **Regra de atualização dos pesos sinápticos da rede neural:** utilizou-se a regra padrão (*default*) de atualização dos pesos sinápticos da *toolbox* de redes neurais do MATLAB.

- Algumas características do processo utilizado de codificação de vértices como vetores:

Com o intuito de se proceder à codificação dos vértices do grafo em vetores, visando à clusterização do mesmo com redes SOM, teve-se que se optar por uma das três estratégias a seguir mencionadas:

Visando-se evitar esparsidade muito elevada nos vetores resultantes do processo de codificação e esforços computacionais desnecessários poder-se-ia utilizar neste processo somente as palavras cujos vértices correspondentes não estavam totalmente isolados (isto é, nesta perspectiva, não seriam codificados os vértices que não possuísem, ligações com nenhum outro, ou seja, que pertenciam a componentes conexas formadas só por eles). Automaticamente se assumiria, que a cada um desses vértices isolados corresponderia um *cluster* diferente.

Tal procedimento proporcionaria os seguintes benefícios:

- Primeiramente, como já antecipado anteriormente, evitar-se-ia esparcidade adicional nos vetores construídos.

- Em segundo lugar, a inclusão, no processo de codificação desses vértices isolados poderia levar à necessidade de se ter que fazer tentativas com diversos valores diferentes para as coordenadas não nulas dos vetores presentes nos componentes adicionais incluídos nos mesmos, em formação (para se produzir o efeito de atribuição de distâncias grandes entre vetores representando vértices em diferentes componentes conexos).

Estas tentativas visariam criar um conjunto de vetores com os quais se pudessem obter um resultado final de clusterização que contivesse cada vetor, representando cada nó isolado do grafo, em um *cluster* isolado, formado só com ele (de forma a garantir que o resultado final da clusterização pudesse atender à prerrogativa essencial, identificada por SCHAEFFER, 2007, a que todo *cluster* em grafos deve obedecer).

Deste modo se poderia poupar todo o esforço simplesmente não se codificando estes vértices isolados do grafo, e conseqüentemente, não os incluindo no treinamento da rede SOM (desde que se entendesse que cada um desses vértices isolados formaria um *cluster* diferente).

Poderia se proceder a uma variação da proposta acima, aplicando-se um processo de codificação independente para cada componente conexo do grafo; e desta forma codificando-se em separado, os vértices pertencentes a cada componente conexa e se treinaria a rede neural SOM para cada uma dessas componentes conexas.

Alternativamente, se incluiria todos os vértices do grafo, sem exceção no processo de codificação.

Optou-se por esta última estratégia por se entender ser útil demonstrar na aplicação proposta, a possibilidade de se utilizar a metodologia de codificação apresentada em grafos cíclicos e desconexos conjuntamente redes SOM (incluindo os vários vértices totalmente isolados). Portanto tal aplicação poderia ilustrar as potencialidades de tal metodologia.

Tais vetores tiveram, como coordenadas não nulas das componentes adicionais incluídas no final do processo de codificação (dos vértices como vetores) valores partindo de 500 (para os vetores representando os vértices do componente conexo 1) e aumentando de unidade em unidade conforme o processo de codificação prosseguiu para vetores representando vértices de outros componentes conexos.

Não houve necessidade de se tentar outros valores para estas coordenadas não nulas dos vetores em questão, uma vez que se obtiveram resultados adequados de clusterização mantendo-se fixa esta codificação e variando-se os demais parâmetros da rede.

O algoritmo de codificação dos vértices do grafo em vetores foi executado em código em linguagem PHP, especificamente desenvolvido para o desenvolvimento deste trabalho.

Estes vetores podem ser encontrados no Apêndice desse trabalho, nas tabelas 1 a 8.

6.2.2 MCL (Markov Clustering)

O algoritmo de Markov Clustering (ou MCL) utilizado neste trabalho foi o disponível em Java, nos sites <http://sourceforge.net/directory/os:windows/freshness:recently-updated/?q=MCL> e <http://micans.org/mcl/>.

Foi implementado a partir do ambiente de desenvolvimento Netbeans IDE versão 7.1.2 64 bits.

O código em Java do algoritmo MCL utilizado possui um critério de parada baseado na energia residual da matriz de Markov atualizada após cada iteração (após o final de cada execução de uma etapa de inflação e de expansão). Este conceito, neste algoritmo, corresponde à diferença entre os elementos das linhas da matriz e a soma dos quadrados dos mesmos. No código Java utilizado pode-se customizar o valor para a energia máxima residual. Por *default* este valor é igual a 0,001; tendo sido mantido neste valor.

Assim, o número de vezes que o algoritmo recorrerá às operações de expansão dependerá do atingimento deste critério de parada e não da definição a priori de número pré-determinado de etapas de expansão.

Com este critério de parada adotado, o algoritmo foi executado para diferentes valores do parâmetro de inflação (em particular foram testados os valores de 2,0; 1,75; 1,4 e 1,85). Os melhores resultados (no sentido de se atender às mencionadas propriedades desejáveis para os clusters no problema analisado) foram obtidos com o parâmetro de inflação igual a 1,85. Esse foi o valor adotado para este parâmetro.

Mais especificamente, com valores de *inflation* muito baixos (como o de 1,4), obtiveram-se *clusters* muito grandes, incorporando muitas palavras distantes entre si, no grafo, ligadas em ordens de relacionamento elevadas. Isto implicava que teria que se aceitarem *clusters* com palavras não muito relacionadas entre si.

Com valores de *inflation* mais altos (como o valor de 2,0), obteve-se numerosos *clusters* muito pequenos (e de elevada granularidade).

O parâmetro que propiciou, no conjunto de dados mencionado, o melhor balanço entre número de *clusters*, tamanho dos *clusters* e principalmente proximidade de sentido entre elas foi o de 1,85.

A identificação destes resultados, para cada diferente parâmetro de *inflation* testado foi por meio do desenho do grafo representante das associações entre estas palavras, e da identificação e destaque, no mesmo, dos *clusters* obtidos com cada tentativa.

CAPÍTULO 7: RESULTADOS DAS CLUSTERIZAÇÕES NO CONJUNTO DE 222 TRAÇOS

A seguir se apresentará os resultados das clusterizações com os dois métodos adotados neste trabalho e se procederá a análise dos mesmos.

7.1 RESULTADOS COM REDES SOM

Analisando-se os resultados obtidos com a implementação das redes neurais de Kohonen observa-se que a rede foi capaz de agrupar palavras próximas (em termos de distancia euclidiana) no TTT nos mesmos *clusters*.

Os *clusters* obtidos por este método estão listados nas tabelas abaixo:

Tabela 25: Resultados obtidos com redes de Kohonen com a partir das 222 *Keywords*

GRUPO CONEXO	Número do <i>Cluster</i>	Traços	Conceitos/ideias gerais associados a cada <i>cluster</i> – SOM
1	<i>Cluster 1</i>	<i>frank, open-hearted</i>	<i>frank</i>
1	<i>Cluster 2</i>	<i>devout, honest, religious, pious, reverent</i>	<i>honest, pious</i>
1	<i>Cluster 3</i>	<i>clever, dexterous, witty</i>	<i>clever</i>
1	<i>Cluster 4</i>	<i>skilful, deft, skilled, prepared, proficient, professional</i>	<i>skilled, profesional</i>
1	<i>Cluster 5</i>	<i>alert, fresh, original</i>	<i>alert, original</i>
1	<i>Cluster 6</i>	<i>economical, frugal, provident, thrifty, sparing</i>	<i>frugal</i>
1	<i>Cluster 7</i>	<i>discriminating, imaginative, aesthetic, polished, informative, hospitable, poised</i>	<i>imaginative, hospitable</i>
1	<i>Cluster 8</i>	<i>social, sociable, gregarious, convivial, genial, cordial, friendly, amiable, affable, political, cooperative, coherent, cosmopolitan, impartial, cheerful, resilient, optimistic, affectionate, kind, responsive, kind-hearted, charming, grateful</i>	<i>gregarious, political, optimistic</i>
42	<i>Cluster 9</i>	<i>pictorial</i>	<i>pictorial</i>
40	<i>Cluster 10</i>	<i>perfidious</i>	<i>perfidious</i>
52	<i>Cluster 11</i>	<i>vindictive</i>	<i>vindictive</i>
2	<i>Cluster 12</i>	<i>fancy</i>	<i>fancy</i>
28	<i>Cluster 13</i>	<i>influencing</i>	<i>influencing</i>
10	<i>Cluster 14</i>	<i>assertive, vocal, speaking, unwritten</i>	<i>assertive</i>
14	<i>Cluster 15</i>	<i>conscientious, painstaking, tactful</i>	<i>conscientious</i>
1	<i>Cluster 16</i>	<i>hearty, generous</i>	<i>hearty</i>
1	<i>Cluster 17</i>	<i>rational, analytical, advisable</i>	<i>analytical</i>

1	Cluster 18	<i>intelligent, logical, wise</i>	<i>intelligent</i>
1	Cluster 19	<i>shrewd, hard</i>	<i>hard</i>
1	Cluster 20	<i>practical, sophisticated, constructive</i>	<i>practical, sophisticated</i>
1	Cluster 21	<i>eclectic, multifarious</i>	<i>eclectic</i>
35	Cluster 22	<i>mischievous</i>	<i>mischievous</i>
9	Cluster 23	<i>argumentative, pugnacious</i>	<i>argumentative</i>
5	Cluster 24	<i>preventive</i>	<i>preventive</i>
38	Cluster 25	<i>observer</i>	<i>observer</i>
23	Cluster 26	<i>governing</i>	<i>governing</i>
54	Cluster 27	<i>wandering</i>	<i>wandering</i>
13	Cluster 28	<i>complaining, cynical, sarcastic</i>	<i>complaining, sarcastic</i>
1	Cluster 29	<i>strong, alcoholic, energetic, courageous, influential</i>	<i>influential, strong</i>
1	Cluster 30	<i>in shape</i>	<i>in shape</i>
1	Cluster 31	<i>active</i>	<i>active</i>
1	Cluster 32	<i>inflexible</i>	<i>inflexible</i>
1	Cluster 33	<i>severe, exact, ambitious, ascetic, serious, enterprising, industrious</i>	<i>severe, industrious</i>
1	Cluster 34	<i>austere, formal</i>	<i>austere</i>
1	Cluster 35	<i>many-sided, multiple</i>	<i>many-sided</i>
18	Cluster 36	<i>easygoing</i>	<i>easygoing</i>
30	Cluster 37	<i>introspective</i>	<i>introspective</i>
7	Cluster 38	<i>acquisitive</i>	<i>acquisitive</i>
24	Cluster 39	<i>habituated</i>	<i>habituated</i>
8	Cluster 40	<i>amorous</i>	<i>amorous</i>
43	Cluster 41	<i>pitying</i>	<i>pitying</i>
41	Cluster 42	<i>perfunctory</i>	<i>perfunctory</i>
45	Cluster 43	<i>reliable, stable</i>	<i>reliable</i>
16	Cluster 44	<i>controlled, patient</i>	<i>controlled</i>
1	Cluster 45	<i>thoughtful, mindful, cautious</i>	<i>thoughtful</i>
1	Cluster 46	<i>visionary, ideal, hypothetical</i>	<i>visionary, hypothetical</i>
1	Cluster 47	<i>reclusive, reserved, inhibited</i>	<i>reclusive</i>
1	Cluster 48	<i>dignified, elevated, eloquent</i>	<i>dignified</i>
1	Cluster 49	<i>conventional</i>	<i>conventional</i>
29	Cluster 50	<i>inhuman</i>	<i>inhuman</i>
32	Cluster 51	<i>laughing</i>	<i>laughing</i>
51	Cluster 52	<i>temperate</i>	<i>temperate</i>
47	Cluster 53	<i>slow</i>	<i>slow</i>
46	Cluster 54	<i>slandorous</i>	<i>slandorous</i>

53	Cluster 55	visual	visual
50	Cluster 56	talkative	talkative
48	Cluster 57	splenetic	splenetic
16	Cluster 58	persevering	persevering
1	Cluster 59	mechanical, programmed	mechanical
1	Cluster 60	observing	observing
1	Cluster 61	doubtful, self-conscious	doubtful, self-conscious
1	Cluster 62	self-satisfied, self-confident	self-confident
27	Cluster 63	independent	independent
11	Cluster 64	athletic	athletic
12	Cluster 65	closed	closed
15	Cluster 66	contented	contented
26	Cluster 67	hurried	hurried
21	Cluster 68	evasive	evasive
33	Cluster 69	leader	leader
1	Cluster 70	standardized	standardized
1	Cluster 71	instinctive, intuitive	instinctive
1	Cluster 72	speculative, conjectural	conjectural
1	Cluster 73	academic	academic
1	Cluster 74	boastful	boastful
1	Cluster 75	conceited, arrogant, egotistical, imitative, autocratic, opinionated, selfish	arrogant, egotistical, opinionated
4	Cluster 76	flattering	flattering
49	Cluster 77	sportive	sportive
19	Cluster 78	eccentric	eccentric
6	Cluster 79	animal	animal
34	Cluster 80	lexical, verbal	verbal
6	Cluster 81	prominent	prominent
1	Cluster 82	impulsive	impulsive
1	Cluster 83	emotional	emotional
1	Cluster 84	idealistic	idealistic
1	Cluster 85	imaginary, subjective	subjective
1	Cluster 86	abstract	abstract
1	Cluster 87	affected	affected
20	Cluster 88	effeminate	effeminate
39	Cluster 89	pedantic	pedantic
3	Cluster 90	fastidious	fastidious
6	Cluster 91	bodily	bodily
6	Cluster 92	physical	physical
6	Cluster 93	visible	visible

1	Cluster 94	<i>vivacious</i>	<i>vivacious</i>
1	Cluster 95	<i>sentimental</i>	<i>sentimental</i>
1	Cluster 96	<i>conceptual</i>	<i>conceptual</i>
1	Cluster 97	<i>worried, frightened</i>	<i>worried</i>
31	Cluster 98	<i>jealous</i>	<i>jealous</i>
25	Cluster 99	<i>headstrong, mulish</i>	<i>headstrong</i>
36	Cluster 100	<i>mystical</i>	<i>mystical</i>
37	Cluster 101	<i>neurotic</i>	<i>neurotic</i>
6	Cluster 102	<i>curious</i>	<i>curious</i>
6	Cluster 103	<i>leading</i>	<i>leading</i>
1	Cluster 104	<i>irritable</i>	<i>irritable</i>
1	Cluster 105	<i>enthusiastic</i>	<i>enthusiastic</i>
1	Cluster 106	<i>excitable</i>	<i>excitable</i>
1	Cluster 107	<i>panicky</i>	<i>panicky</i>
1	Cluster 108	<i>timid</i>	<i>timid</i>
17	Cluster 109	<i>decorative</i>	<i>decorative</i>
22	Cluster 110	<i>gluttonous</i>	<i>gluttonous</i>
44	Cluster 111	<i>plaintive</i>	<i>plaintive</i>
6	Cluster 112	<i>corporal</i>	<i>corporal</i>
6	Cluster 113	<i>interested</i>	<i>interested</i>
1	Cluster 114	<i>sour</i>	<i>sour</i>
1	Cluster 115	<i>sensitive</i>	<i>sensitive</i>
1	Cluster 116	<i>high-strung</i>	<i>high-strung</i>
1	Cluster 117	<i>anxious</i>	<i>anxious</i>
1	Cluster 118	<i>afraid</i>	<i>afraid</i>
1	Cluster 119	<i>coward</i>	<i>coward</i>
17	Cluster 120	<i>musical</i>	<i>musical</i>
6	Cluster 121	<i>corporeal</i>	<i>corporeal</i>
6	Cluster 122	<i>perceivable</i>	<i>perceivable</i>

*Células em vermelho representam componentes conexos formados por vértices (traços) isolados.

*Células em amarelo representam palavras que pertenciam a componentes conexos com mais de um elemento, mas que ficaram isoladas na clusterização.

Estes *clusters* foram obtidos pela alocação dos vetores nas diversas regiões do mapa topológico.

Este mapa é apresentado abaixo, conjuntamente com as regiões do mesmo que receberam vetores:

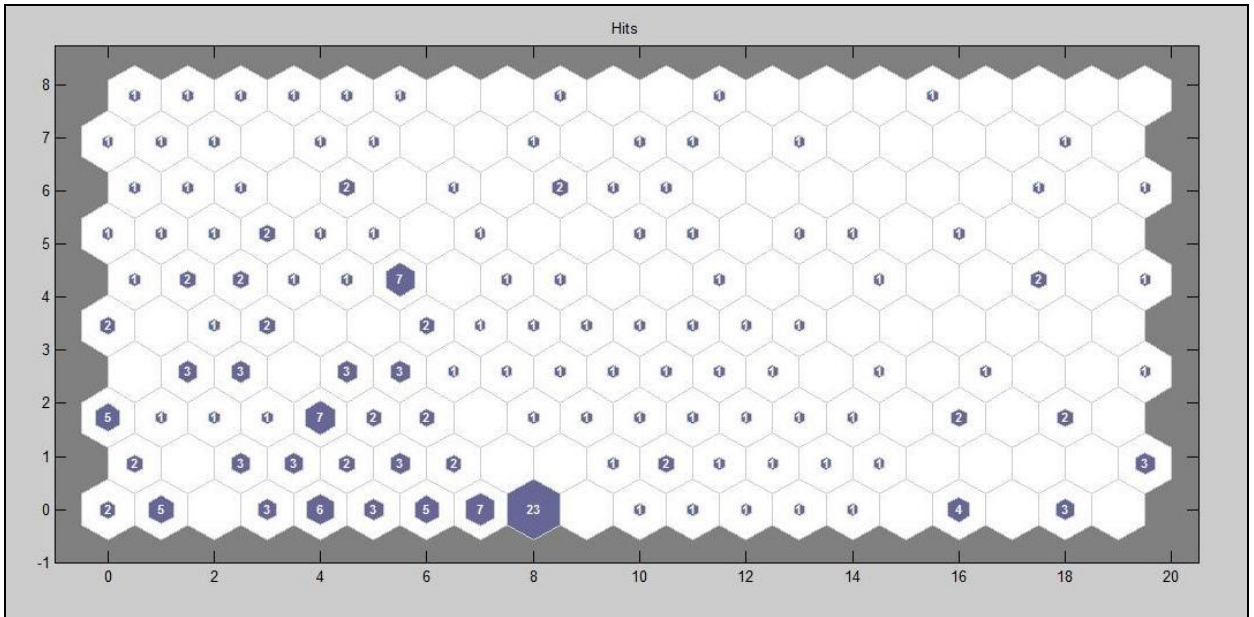


Figura 32: Alocação final no mapa topológico utilizado na rede SOM

Fonte: Relatório *Sample Hits* da *Toolbox Neural Networks* do MATLAB.

As palavras correspondentes a cada região do mapa acima, podem ser identificadas na tabela abaixo.

Tabela 26: Alocação das palavras nas regiões do mapa topológico.

GRUPO CONEXO	TRAÇO	REGIÃO DO MAPA ONDE FOI ALOCADA
1	<i>frank</i>	1
1	<i>open-hearted</i>	1
1	<i>devout</i>	2
1	<i>honest</i>	2
1	<i>religious</i>	2
1	<i>pious</i>	2
1	<i>reverent</i>	2
1	<i>clever</i>	4
1	<i>dexterous</i>	4
1	<i>witty</i>	4
1	<i>skilful</i>	5
1	<i>deft</i>	5
1	<i>skilled</i>	5
1	<i>prepared</i>	5
1	<i>proficient</i>	5
1	<i>professional</i>	5
1	<i>alert</i>	6
1	<i>fresh</i>	6

1	<i>original</i>	6
1	<i>economical</i>	7
1	<i>frugal</i>	7
1	<i>provident</i>	7
1	<i>thrifty</i>	7
1	<i>sparing</i>	7
1	<i>discriminating</i>	8
1	<i>imaginative</i>	8
1	<i>aesthetic</i>	8
1	<i>polished</i>	8
1	<i>informative</i>	8
1	<i>hospitable</i>	8
1	<i>poised</i>	8
1	<i>social</i>	9
1	<i>sociable</i>	9
1	<i>gregarious</i>	9
1	<i>convivial</i>	9
1	<i>genial</i>	9
1	<i>cordial</i>	9
1	<i>friendly</i>	9
1	<i>amiable</i>	9
1	<i>affable</i>	9
1	<i>political</i>	9
1	<i>cooperative</i>	9
1	<i>coherent</i>	9
1	<i>cosmopolitan</i>	9
1	<i>impartial</i>	9
1	<i>cheerful</i>	9
1	<i>resilient</i>	9
1	<i>optimistic</i>	9
1	<i>affectionate</i>	9
1	<i>kind</i>	9
1	<i>responsive</i>	9
1	<i>kind-hearted</i>	9
1	<i>charming</i>	9
1	<i>grateful</i>	9
42	<i>pictorial</i>	11
40	<i>perfidious</i>	12
52	<i>vindictive</i>	13
2	<i>fancy</i>	14
28	<i>influencing</i>	15

10	<i>assertive</i>	17
10	<i>vocal</i>	17
10	<i>speaking</i>	17
10	<i>unwritten</i>	17
14	<i>conscientious</i>	19
14	<i>painstaking</i>	19
14	<i>tactful</i>	19
1	<i>hearty</i>	21
1	<i>generous</i>	21
1	<i>rational</i>	23
1	<i>analytical</i>	23
1	<i>advisable</i>	23
1	<i>intelligent</i>	24
1	<i>logical</i>	24
1	<i>wise</i>	24
1	<i>shrewd</i>	25
1	<i>hard</i>	25
1	<i>practical</i>	26
1	<i>sophisticated</i>	26
1	<i>constructive</i>	26
1	<i>eclectic</i>	27
1	<i>multifarious</i>	27
35	<i>mischievous</i>	30
9	<i>argumentative</i>	31
9	<i>pugnacious</i>	31
5	<i>preventive</i>	32
38	<i>observer</i>	33
23	<i>governing</i>	34
54	<i>wandering</i>	35
13	<i>complaining</i>	40
13	<i>cynical</i>	40
13	<i>sarcastic</i>	40
1	<i>strong</i>	41
1	<i>alcoholic</i>	41
1	<i>energetic</i>	41
1	<i>courageous</i>	41
1	<i>influential</i>	41
1	<i>in shape</i>	42
1	<i>active</i>	43
1	<i>inflexible</i>	44
1	<i>severe</i>	45

1	<i>exact</i>	45
1	<i>ambitious</i>	45
1	<i>ascetic</i>	45
1	<i>serious</i>	45
1	<i>enterprising</i>	45
1	<i>industrious</i>	45
1	<i>austere</i>	46
1	<i>formal</i>	46
1	<i>many-sided</i>	47
1	<i>multiple</i>	47
18	<i>easygoing</i>	49
30	<i>introspective</i>	50
7	<i>acquisitive</i>	51
24	<i>habituated</i>	52
8	<i>amorous</i>	53
43	<i>pitying</i>	54
41	<i>perfunctory</i>	55
45	<i>reliable</i>	57
45	<i>stable</i>	57
16	<i>controlled</i>	59
16	<i>patient</i>	59
1	<i>thoughtful</i>	62
1	<i>mindful</i>	62
1	<i>cautious</i>	62
1	<i>visionary</i>	63
1	<i>ideal</i>	63
1	<i>hypothetical</i>	63
1	<i>reclusive</i>	65
1	<i>reserved</i>	65
1	<i>inhibited</i>	65
1	<i>dignified</i>	66
1	<i>elevated</i>	66
1	<i>eloquent</i>	66
1	<i>conventional</i>	67
29	<i>inhuman</i>	68
32	<i>laughing</i>	69
51	<i>temperate</i>	70
47	<i>slow</i>	71
46	<i>slanderous</i>	72
53	<i>visual</i>	73
50	<i>talkative</i>	75

48	<i>splenic</i>	77
16	<i>persevering</i>	80
1	<i>mechanical</i>	81
1	<i>programmed</i>	81
1	<i>observing</i>	83
1	<i>doubtful</i>	84
1	<i>self-conscious</i>	84
1	<i>self-satisfied</i>	87
1	<i>self-confident</i>	87
27	<i>independent</i>	88
11	<i>athletic</i>	89
12	<i>closed</i>	90
15	<i>contented</i>	91
26	<i>hurried</i>	92
21	<i>evasive</i>	93
33	<i>leader</i>	94
1	<i>standardized</i>	101
1	<i>instinctive</i>	102
1	<i>intuitive</i>	102
1	<i>speculative</i>	103
1	<i>conjectural</i>	103
1	<i>academic</i>	104
1	<i>boastful</i>	105
1	<i>conceited</i>	106
1	<i>arrogant</i>	106
1	<i>egotistical</i>	106
1	<i>imitative</i>	106
1	<i>autocratic</i>	106
1	<i>opinionated</i>	106
1	<i>selfish</i>	106
4	<i>flattering</i>	108
49	<i>sportive</i>	109
19	<i>eccentric</i>	112
6	<i>animal</i>	115
34	<i>lexical</i>	118
34	<i>verbal</i>	118
6	<i>prominent</i>	120
1	<i>impulsive</i>	121
1	<i>emotional</i>	122
1	<i>idealistic</i>	123
1	<i>imaginary</i>	124

1	<i>subjective</i>	124
1	<i>abstract</i>	125
1	<i>affected</i>	126
20	<i>effeminate</i>	128
39	<i>pedantic</i>	131
3	<i>fastidious</i>	132
6	<i>bodily</i>	134
6	<i>physical</i>	135
6	<i>visible</i>	137
1	<i>vivacious</i>	141
1	<i>sentimental</i>	142
1	<i>conceptual</i>	143
1	<i>worried</i>	145
1	<i>frightened</i>	145
31	<i>jealous</i>	147
25	<i>headstrong</i>	149
25	<i>mulish</i>	149
36	<i>mystical</i>	150
37	<i>neurotic</i>	151
6	<i>curious</i>	158
6	<i>leading</i>	160
1	<i>irritable</i>	161
1	<i>enthusiastic</i>	162
1	<i>excitable</i>	163
1	<i>panicky</i>	165
1	<i>timid</i>	166
17	<i>decorative</i>	169
22	<i>gluttonous</i>	171
44	<i>plaintive</i>	172
6	<i>corporal</i>	174
6	<i>interested</i>	179
1	<i>sour</i>	181
1	<i>sensitive</i>	182
1	<i>high-strung</i>	183
1	<i>anxious</i>	184
1	<i>afraid</i>	185
1	<i>coward</i>	186
17	<i>musical</i>	189
6	<i>corporeal</i>	192
6	<i>perceivable</i>	196

Um aspecto importante a ser ressaltado nos resultados obtidos por SOM é que, em boa parte dos casos os *clusters* encontrados pela rede foram formados por palavras imediatamente vizinhas ou relacionadas entre si em até segunda ordem (ou seja, com distância euclidiana igual a dois). Como visto anteriormente, em um *thesaurus* palavras vizinhas são em sua maioria sinônimas entre si.

Este aspecto é desejável do ponto da organização de palavras semelhantes semanticamente nos mesmos subconjuntos. Adicionalmente, com estes resultados também se verifica que os sentidos gerais dos *clusters* encontrados, também tendem a ser mais próximos entre si (para os *clusters* onde se identificou mais de um sentido geral sintetizando as ideias dos traços neles contidos).

Outro aspecto de extrema relevância e que deve ser ressaltado de imediato é que, embora se esteja ciente do fato de que a codificação proposta para as palavras do TTT (como vetores de alta dimensão) foi idealizada para se atribuir distâncias entre vértices de grafos acíclicos, redes SOM foram usadas (utilizando como *input* tais vetores) no grafo correspondente às relações entre as 222 palavras selecionadas.

Os motivos para esta utilização foram os seguintes:

- O problema abordado neste trabalho se limitava a poucas centenas de palavras (e, portanto os sintomas decorrentes da maldição da dimensionalidade não teriam influência relevante; de modo que as distâncias computadas no treinamento da rede não sofreram significativa influência deste fenômeno).

- Em função de, embora o algoritmo de codificação de vértices proposto forneça apenas aproximações para as distâncias entre um determinado par de vértices para cada ciclo do grafo (funcionando como um método de aproximação destas distâncias, para estes pares específicos de vértices) deve-se observar que:

a) No problema ora abordado os ciclos existentes neste grafo são poucos e curtos (de modo que, em cada ciclo identificado, se estaria errando pouco na atribuição de distâncias para algumas palavras em cada ciclo).

b) O treinamento da rede SOM computa as distâncias entre vetores de entrada e os modelos que a rede constrói para representar estes dados, e não propriamente a distância entre os vetores de entrada (embora durante o treinamento e aprendizagem da rede as distâncias entre estes vetores são refletidas indiretamente na alocação dos mesmos em *clusters*, de modo que estas distâncias importam e exercem influência na alocação final dos vetores em *clusters*, pela rede).

- A alocação de palavras pela rede SOM é função não só das distâncias computadas entre os elementos de entrada e os modelos que a rede constrói para representá-los, mas também de outras características definidas para a rede e para o processo de treinamento, tais como:

- a) a arquitetura do mapa topológico da rede,
- b) o número de épocas a ser apresentado para a rede SOM durante o treinamento,
- c) a forma como as vizinhanças serão definidas no entorno de cada modelo e como seus raios serão atualizados ao longo do treinamento e
- d) a taxa de aprendizagem utilizada no treinamento.

Um algoritmo de clusterização que utilizasse somente as distâncias entre os vetores assim construídos iria, em um grafo cíclico, alocar em *clusters* diferentes palavras pertencentes a um mesmo ciclo (mais especificamente, para cada ciclo existente no grafo, uma palavra pertencente ao mesmo poderia ser alocada em um *cluster* em separado das demais). No entanto, na rede SOM a alocação final de objetos em *clusters* depende também desses outros fatores.

Neste sentido, não se observaram alocações feitas pela rede, de vértices originalmente pertencentes a ciclos, onde se verificasse tal fenômeno (isto é, a identificação de vértices pertencentes a um mesmo ciclo onde um desses vértices seja alocado em um determinado *cluster* e todos os demais pertencentes ao mesmo ciclo, em outro), de modo que, julgando-se pelos resultados obtidos, o algoritmo proposto de codificação de vértices em vetores produziu bons resultados para o problema em questão.

Ilustrando com a qualidade da alocação das palavras feita pela rede, deve-se ressaltar que a rede SOM foi capaz de alocar todos os vértices pertencentes à única clique do grafo (formada pelos traços *thrifty*, *frugal*, *sparing*, *economical*, e *provident*) em um único *cluster* (número 6).

Como a cada região (hexagonal) do mapa topológico da rede neural que ficou com elementos ao final do treinamento, corresponde a um *cluster*, pode-se notar na tabela acima que esta rede definiu 122 *clusters* a partir das 222 palavras utilizadas, com o esquema de codificação proposto para os nós do grafo como vetores. Desses, 38 possuem pelo menos dois elementos (traços de personalidade) e os demais são constituídos apenas por um único traço.

Esses 38 *clusters* contém a grande parte das palavras do TTT (138 palavras, correspondendo a 62,16 % das palavras selecionadas), as demais 84 palavras foram alocadas em *clusters* contendo somente elas (formando-se assim 84 *clusters*).

Desses 84 *clusters*, 43 correspondem a traços que já estavam totalmente isolados no TTT e, portanto deveriam ser alocados em *clusters* dedicados. Deste modo é correto afirmar que, das 179 palavras cujos vértices pertenciam a componentes conexos formados por mais de um elemento, somente 41 ficaram isoladas na alocação final da rede (representando 18,47 % do universo de 222 traços empregados). Pode-se verificar assim que a clusterização obtida por esta metodologia não ultrapassou o limite dado a priori de 20%, dado para que palavras com tais características fossem alocadas em *clusters* isolados.

Todos os demais *clusters* definidos pela rede, com as configurações adotadas são compostos por mais de um elemento.

Adicionalmente, pode-se notar inicialmente que os *clusters* encontrados pela rede neural foram compostos de palavras que de fato compartilham um (alguns) conceito(s) - ideia(s) – mais amplo(s), comum(s) às palavras neles contidos. Estes sentidos mais amplos encontram-se listados na tabela 21 acima.

Como já ressaltado anteriormente, esta(s) ideia (as) geral (is) representativa(s) de cada *cluster* não necessariamente reproduz (em) com exatidão todos os significados e interpretações possíveis que cada palavra isoladamente, em cada *cluster*, pode assumir. Entretanto esta(s) ideia (as) geral (is) fornece uma descrição geral, uma *rich picture*, dos conceitos contidos em cada um destes *clusters*.

Outro ponto que de imediato merece ser destacado é que a rede, conjuntamente com a adoção da metodologia de codificação dos vértices do grafo como vetores foram capazes de produzir *clusters* contendo elementos conectados entre si. Tal prerrogativa é indispensável na caracterização de tais agrupamentos; haja vista que esta é uma característica indispensável (como já abordado anteriormente) a que todo *cluster* em grafos deve obedecer para ser caracterizado como tal.

Logicamente, esta exigência demandaria que os vértices do grafo que originalmente já estavam completamente isolados dos demais formassem cada um, um *cluster* distinto; o que de fato também se verificou.

A maioria dos *clusters* encontrados com esta metodologia contém dois ou três elementos.

O maior dos componentes conexos do grafo (componente 1), contendo 145 vértices (ou traços) dos 222 selecionados foi particionado em 56 *clusters* (rotulados com os números 1 a 8, 16 a 21, 29 a 35, 45 a 49, 59 a 62, 70 a 75, 82 a 87, 94 a 97, 104 a 108, 114 a 119).

O componente conexo 6 foi particionado em 11 *clusters*; rotulados pelos números 79, 81, 91, 92, 93, 102, 103, 112, 113, 121 e 122. Do mesmo modo, o componente conexo 17, também foi particionado em dois *clusters* (109 e 120).

Analisando mais detalhadamente os *clusters* formados por mais de uma palavra, encontrados por este método, tem-se:

- No *cluster* 1 (composto por *Frank* e *open-hearted*) verificaram-se palavras relacionadas às ideias de ser franco e de coração aberto. Assim, escolheu-se a palavra *frank* (franco) para representar as ideias gerais deste agrupamento.

- *Cluster* 2: formado por *devout, honest, religious, pious, reverent*. Este *cluster* está associado às ideias de ser honesto, devoto e religioso. Assim, escolheram-se as palavras *honest* e *pious* para representarem as ideias gerais desse conjunto.

- *Cluster* 3: formado por *clever, dexterous e witty*. Este *cluster* está associado às ideias de ser honesto, esperto e habilidoso. Assim, escolheu-se a palavra *clever* para representar as ideias gerais desse conjunto.

- *Cluster* 4: composto por *skilful, deft, skilled, prepared, proficient e professional*. Este *cluster* está associado às ideias de ser preparado, profissional e ter habilidades. Assim, escolheu-se as palavras *skilled* e *professional* para representar as ideias gerais desse conjunto.

- *Cluster* 5: contendo as palavras *alert, fresh e original*. Este *cluster* está associado às ideias de ser alerta e original. Assim, escolheu-se *alert* e *original* para representar as ideias gerais desse conjunto.

- *Cluster* 6: contendo a clique composta *economical, frugal, provident, thrifty e sparing*. Este *cluster* está associado às ideias de ser frugal, econômico. Assim, escolheu-se a palavra *frugal* para representar as ideias gerais desse conjunto.

- *Cluster* 7: formado por *discriminating, imaginative, aesthetic, polished, informative, hospitable, poised*. Escolheu-se as palavras *imaginative e hospitable* para representar as ideias gerais desse conjunto.

- *Cluster* 8: formado por *social, sociable, gregarious, convivial, genial, cordial, friendly, amiable, affable, political, cooperative, coherent, cosmopolitan, impartial, cheerful, resilient, optimistic, affectionate, kind, responsive, kind-hearted, charming e grateful*.

Escolheram-se as palavras *gregarious*, *political* e *optimistic* para representarem as ideias gerais desse conjunto.

- *Cluster 14*: formado por *assertive*, *vocal*, *speaking* e *unwritten*. Ligado às ideias de ser assertivo, vocal e falante. Escolheu-se a palavra *assertive* para representar as ideias gerais desse conjunto.

- *Cluster 15*: formado por *conscientious*, *painstaking* e *tactful*. Ligado às ideias de ser consciente e se ter tato, ser hábil. Escolheu-se a palavra *conscientious* para representar as ideias gerais desse conjunto.

- *Cluster 16*: contém os termos *hearty* e *generous*. Ligado às ideias de se ter coração e ser generoso. Escolheu-se a palavra *hearty* para representar as ideias gerais desse conjunto.

- *Cluster 17*: contém os termos *rational*, *analytical* e *advisable*. Ligado às ideias de ser racional, analítico e de ser aconselhável, recomendável. Escolheu-se a palavra *analytical* para representar as ideias gerais desse conjunto.

- *Cluster 18*: contemplando os termos *intelligent*, *logical* e *wise*. Ligado às ideias de ser inteligente, lógico e sábio. Escolheu-se a palavra *intelligent* para representar as ideias gerais desse conjunto.

- *Cluster 19*: contemplando os termos *shrewd* e *hard*. Ligado às ideias de ser astuto, agudo, sagaz e também de ser duro, resistente e forte. As duas palavras foram utilizadas para representar as ideias gerais desse conjunto.

- *Cluster 20*: contempla as palavras *practical*, *sophisticated* e *constructive*. Ligado às ideias de ser prático, sofisticado e construtivo. As palavras *practical* e *sophisticated* foram utilizadas para representar as ideias gerais desse conjunto.

- *Cluster 21*: contempla as palavras *eclectic* e *multifarious*. Ligado às ideias de ser eclético e multifário, com múltiplas habilidade e facetas. Escolheu-se *eclectic* para representar as ideias gerais desse conjunto.

- *Cluster 23*: contempla as palavras *argumentative* e *pugnacious*. Ligado às ideias de ser argumentativo e belicoso. Escolheu-se *argumentative* para representar as ideias gerais desse conjunto.

- *Cluster 28*: contempla as palavras *complaining*, *cynical* e *sarcastic*. Ligado às ideias de ser alguém que reclama constantemente, de ser cínico e sarcástico. Escolheram-se os termos *complaining* e *sarcastic* para representarem estes conceitos.

- *Cluster 29*: este agrupamento contempla as palavras *strong*, *alcoholic*, *energetic*, *courageous* e *influential*. Remete às ideias de se ser influente e com personalidade forte, que

influencia ou que marca os demais. Escolheram-se os termos *influential* e *strong* para representarem estes conceitos.

- *Cluster 33*: são formados por *severe*, *exact*, *ambitious*, *ascetic*, *serious*, *enterprising* e empreendedor. Escolheram-se os termos *severe* e *industrious* para representarem estes conceitos.

- *Cluster 34*: formado por *austere* e *formal*. Compreende as ideias de se ser *austero* e *formal*. Escolheu-se o termo *austere* para representar este *cluster*.

- *Cluster 35*: formado por *many-sided* e *multiple*. Compreende as ideias de se ser *multifacetado*. Escolheu-se o termo *many-sided* para representar este *cluster*.

- *Cluster 43*: formado por *many-sided* e *multiple*. Compreende as ideias de se ser *multifacetado*. Escolheu-se o termo *many-sided* para representar este *cluster*.

- *Cluster 44*: formado por *controlled* e *patient*. Compreende as ideias de se ser controlado e paciente. Escolheu-se o termo *controlled* para representar este *cluster*.

- *Cluster 45*: formado por *thoughtful*, *mindful* e *cautious*. Compreende as ideias de se ser pensativo e reflexivo. Escolheu-se o termo *thoughtful* para representar este *cluster*.

- *Cluster 46*: formado por *visionary*, *ideal* e *hypothetical*. Compreende as ideias de se ser visionário, ideal (idealista) e hipotético. Escolheram-se os termos *visionary* e *hypothetical* para representarem tais ideias.

- *Cluster 47*: formado por *reclusive*, *reserved* e *inhibited*. Compreende as ideias de se ser reclusivo, reservado e inibido. Escolheu-se o termo *reclusive* para representar essas ideias.

- *Cluster 48*: formado por *dignified*, *elevated* e *eloquent*. Compreende as ideias de se ser dignificado, digno, elevado e eloquente. Escolheu-se o termo *dignified* para representar essas ideias.

- *Cluster 61*: formado por *doubtful* e *self-conscious*. Compreende as ideias de se ser duvidoso, que gosta de discutir/argumentar e consciente de si mesmo. Por este agrupamento conter somente essas duas palavras (e de sentidos diferentes), ambas as palavras foram empregadas para representar as ideias desse conjunto.

- *Cluster 62*: formado por *self-satisfied* e *self-confident*. Compreende as ideias de se ser auto-satisfeito, satisfeito consigo mesmo e auto-confiante. Escolheu-se o termo *self-confident* para representar estas ideias.

- *Cluster 71*: formado por *instinctive* e *intuitive*. Compreende as ideias de se ser instintivo e intuitivo. Escolheu-se o termo *instinctive* para representar estas ideias.

- *Cluster 75*: englobando os termos *conceited, arrogant, egotistical, imitative, autocratic, opinionated, selfish*. Engloba as ideias de se ser vaidoso, arrogante, egoísta, teimoso e imitativo. Escolheram-se os termos *arrogant, egotistical* e *opinionated* para representarem estas ideias.

- *Cluster 80*: englobando os termos *lexical* e *verbal*. Engloba as ideias de se ser léxico e verbal. Escolheu-se o termo *verbal* para representar estas ideias.

- *Cluster 85*: englobando os termos *imaginary* e *subjective*. Engloba as ideias de se ser imaginário e subjetivo. Escolheu-se o termo *subjective* para representar estas ideias.

- *Cluster 97*: englobando os termos *worried* e *frightened*. Engloba as ideias de se ser preocupado e espantado. Escolheu-se o termo *worried* para representar estas ideias.

- *Cluster 99*: englobando os termos *headstrong* e *mulish*. Engloba as ideias de se ser cabeça dura e teimoso. Escolheu-se o termo *headstrong* para representar estas ideias.

Todos os demais *clusters* são compostos por um único traço.

Deste modo, pode-se averiguar que a rede neural de Kohonen, conjuntamente com a metodologia de codificação dos vértices do grafo representando as relações entre os traços selecionados para esta aplicação foram capazes de agrupar tais traços em conjuntos (*clusters*) onde os mesmos apresentaram grande similaridade no que se refere aos aspectos comportamentais por eles representados; o que é desejável do ponto de vista de uma ferramenta de classificação e organização dos mesmos.

Também pode ser verificado que boa parte dos *clusters* encontrados apresentou um alto grau de conectividade entre suas palavras e menor conectividade inter *cluster*. Como visto anteriormente esta característica também é desejável no problema abordado.

As figuras 36 a 97 no Apêndice apresentam os *clusters* identificados nas diversas componentes conexas do grafo, formados por mais de um traço.

Outra análise realizada nos resultados encontrados com redes SOM foi no que diz respeito às ligações entre os vértices do grafo e o percentual de ligações entre os mesmos dentro de um mesmo *cluster* (ligações *intra clusters*) e entre *clusters* diferentes (*inter clusters*).

Esta análise é relevante pois pode-se impor como restrição adicional para que um determinado resultado de clusterização seja aceito, o fato de que a maior parte das ligações entre vértices seja *intra cluster*.

Neste sentido, as tabelas 27 e 28 abaixo apresentam estas informações para os *clusters* encontrados por redes SOM.

Tabela 27: Percentuais de ligações *intra* e *inter clusters* verificados para os resultados obtidos com o algoritmo MCL na componente conexa 1 do grafo.

Resultados SOM								
Compo. Conexa do grafo	Cluster #	Rótulos dos <i>clusters</i>	# traços	Total de ligações dos traços do <i>cluster</i>	# de ligações <i>intra cluster</i>	# de ligações <i>inter clusters</i>	% de ligações <i>intra clusters</i>	% de ligações <i>inter clusters</i>
1	1	<i>frank</i>	2	2	1	1	50,00%	50,00%
1	2	<i>honest, pious</i>	5	7	5	2	71,43%	28,57%
1	3	<i>clever</i>	3	8	2	6	25,00%	75,00%
1	4	<i>skilled, professional</i>	6	10	7	3	70,00%	30,00%
1	5	<i>alert, original</i>	3	7	2	5	28,57%	71,43%
1	6	<i>frugal</i>	5	10	9	1	90,00%	10,00%
1	7	<i>imaginative, hospitable</i>	7	14	8	6	57,14%	42,86%
1	8	<i>gregarious, political, optimistic</i>	23	42	35	7	83,33%	16,67%
1	16	<i>heartly</i>	2	11	1	10	9,09%	90,91%
1	17	<i>analytical</i>	3	6	2	4	33,33%	66,67%
1	18	<i>intelligent</i>	3	9	2	7	22,22%	77,78%
1	19	<i>hard</i>	2	9	1	8	11,11%	88,89%
1	20	<i>practical, sophisticated</i>	3	6	2	4	33,33%	66,67%
1	21	<i>eclectic</i>	2	4	1	3	25,00%	75,00%
1	29	<i>influential, strong</i>	5	9	4	5	44,44%	55,56%
1	30	<i>in shape</i>	1	1	0	1	0,00%	100,00%
1	31	<i>active</i>	1	1	0	1	0,00%	100,00%
1	32	<i>inflexible</i>	1	3	0	3	0,00%	100,00%
1	33	<i>severe, industrious</i>	7	15	6	9	40,00%	60,00%
1	34	<i>austere</i>	2	9	1	8	11,11%	88,89%
1	35	<i>many-sided</i>	2	3	1	2	33,33%	66,67%
1	45	<i>thoughtful</i>	3	4	2	2	50,00%	50,00%
1	46	<i>visionary, hypothetical</i>	3	4	2	2	50,00%	50,00%

1	47	<i>reclusive</i>	3	5	3	2	60,00%	40,00%
1	48	<i>dignified</i>	3	4	2	2	50,00%	50,00%
1	49	<i>conventional</i>	1	1	0	1	0,00%	100,00%
1	59	<i>mechanical</i>	2	3	1	2	33,33%	66,67%
1	60	<i>observing</i>	1	3	0	3	0,00%	100,00%
1	61	<i>doubtful, self-conscious</i>	2	2	1	1	50,00%	50,00%
1	62	<i>self-confident</i>	2	4	1	3	25,00%	75,00%
1	70	<i>standardized</i>	1	1	0	1	0,00%	100,00%
1	71	<i>instinctive</i>	2	3	1	2	33,33%	66,67%
1	72	<i>conjectural</i>	2	4	1	3	25,00%	75,00%
1	73	<i>academic</i>	1	1	0	1	0,00%	100,00%
1	74	<i>boastful</i>	1	2	0	2	0,00%	100,00%
1	75	<i>arrogant, egotistical, opinionated</i>	7	13	7	6	53,85%	46,15%
1	82	<i>impulsive</i>	1	3	0	3	0,00%	100,00%
1	83	<i>emotional</i>	1	2	0	2	0,00%	100,00%
1	84	<i>idealistic</i>	1	2	0	2	0,00%	100,00%
1	85	<i>subjective</i>	2	2	1	1	50,00%	50,00%
1	86	<i>abstract</i>	1	3	0	3	0,00%	100,00%
1	87	<i>affected</i>	1	1	0	1	0,00%	100,00%
1	94	<i>vivacious</i>	1	2	0	2	0,00%	100,00%
1	95	<i>sentimental</i>	1	3	0	3	0,00%	100,00%
1	96	<i>conceptual</i>	1	3	0	3	0,00%	100,00%
1	97	<i>worried</i>	2	2	1	1	50,00%	50,00%
1	104	<i>irritable</i>	1	4	0	4	0,00%	100,00%
1	105	<i>enthusiastic</i>	1	5	0	5	0,00%	100,00%
1	106	<i>excitable</i>	1	3	0	3	0,00%	100,00%
1	107	<i>panicky</i>	1	2	0	2	0,00%	100,00%
1	108	<i>timid</i>	1	2	0	2	0,00%	100,00%
1	114	<i>sour</i>	1	2	0	2	0,00%	100,00%
1	115	<i>sensitive</i>	1	1	0	1	0,00%	100,00%
1	116	<i>high-strung</i>	1	1	0	1	0,00%	100,00%
1	117	<i>anxious</i>	1	4	0	4	0,00%	100,00%
1	118	<i>afraid</i>	1	3	0	3	0,00%	100,00%
1	119	<i>coward</i>	1	1	0	1	0,00%	100,00%
Valores médios na comp. 1:			5,11	1,98	3,12	22,26%	77,74%	

Nota: na tabela acima não foram considerados os traços pertencentes a componentes conexas originalmente compostas somente por um único termo.

Tabela 28: Percentuais de ligações *intra* e *inter clusters* verificados para os resultados obtidos com o algoritmo MCL nas demais componentes conexas do grafo.

Resultados SOM								
Compo. Conexa do grafo	Cluster #	Rótulos dos clusters	# traços	Total de ligações dos traços do cluster	# de ligações <i>intra cluster</i>	# de ligações <i>inter clusters</i>	% de ligações <i>intra clusters</i>	% de ligações <i>inter clusters</i>
6	79	<i>animal</i>	1	1	0	1	0,00%	100,00%
6	81	<i>prominent</i>	1	2	0	2	0,00%	100,00%
6	91	<i>bodily</i>	1	3	0	3	0,00%	100,00%
6	92	<i>physical</i>	1	2	0	2	0,00%	100,00%
6	93	<i>visible</i>	1	2	0	2	0,00%	100,00%
6	102	<i>curious</i>	1	3	0	3	0,00%	100,00%
6	103	<i>leading</i>	1	1	0	1	0,00%	100,00%
6	112	<i>corporal</i>	1	3	0	3	0,00%	100,00%
6	113	<i>interested</i>	1	1	0	1	0,00%	100,00%
6	121	<i>corporeal</i>	1	4	0	4	0,00%	100,00%
6	122	<i>perceivable</i>	1	1	0	1	0,00%	100,00%
9	23	<i>argumentative</i>	2	1	1	0	100,00%	0,00%
13	28	<i>complaining, sarcastic</i>	3	2	2	0	100,00%	0,00%
16	44	<i>controlled</i>	3	2	2	0	100,00%	0,00%
16	58	<i>persevering</i>	3	2	2	0	100,00%	0,00%
17	109	<i>decorative</i>	2	1	1	0	100,00%	0,00%
17	120	<i>musical</i>	2	1	1	0	100,00%	0,00%
25	99	<i>headstrong</i>	2	1	1	0	100,00%	0,00%
34	80	<i>verbal</i>	2	1	1	0	100,00%	0,00%
45	43	<i>reliable</i>	2	1	1	0	100,00%	0,00%

Nota: na tabela acima não foram considerados os traços pertencentes a componentes conexas originalmente compostas somente por um único termo.

Como se pode observar nas tabelas 27 e 28 acima, com as escolhas para os parâmetros da rede SOM, para a arquitetura de seu mapa topológico e para a estratégia de codificação dos vértices do grafo associado à rede de ligações entre os traços considerados, os *clusters* encontrados foram compostos por poucas palavras em relação aos encontrados por MCL.

Como consequência, nos vértices do grafo predominaram ligações *inter clusters* em relação às *intra clusters* (para os vértices da componente conexa 1 do grafo, 77,74% das ligações entre vértices ocorreram entre vértices pertencentes a *clusters* diferentes).

Caso se desejasse impor, a restrição de que a maior parte das ligações entre vértices seja *intra cluster*, seria necessário se promover alterações nos itens citados.

7.2 RESULTADOS COM MCL (MARKOV CLUSTERING)

Com a escolha do parâmetro de *inflation* no valor de 1,85 (valor escolhido), obteve-se o seguinte resultado apresentado na figura abaixo:

Estes *clusters* estão representados e organizados na tabela abaixo:

Tabela 22: *Clusters* obtidos com o algoritmo MCL com parâmetro de *inflation* = 1,85.

Resultados com MCL. Inflation = 1,85			
GRUPO CONEXO	<i>Clusters</i>	Palavras	Conceitos/idéias gerais associados a cada <i>cluster</i> - MCL
1	<i>Cluster 1</i>	<i>intelligent, advisable, alert, logical, shrewd, witty, proficient, clever, skilled, wise, dexterous, professional, active, prepared, skilful, analytical, rational, observing, deft</i>	<i>clever, analytical</i>
1	<i>Cluster 2</i>	<i>discriminating, imaginative, multifarious, eclectic, many-sided, multiple</i>	<i>eclectic</i>
1	<i>Cluster 3</i>	<i>cosmopolitan, social, cordial, hospitable, cheerful, gregarious, poised, kind, aesthetic, coherent, kind-hearted, friendly, affable, affectionate, charming, polished, sociable, grateful, political, amiable, convivial, impartial, genial, cooperative, responsive, informative</i>	<i>sociable, amiable</i>
1	<i>Cluster 4</i>	<i>programmed, mechanical, standardized, instinctive, intuitive</i>	<i>instinctive</i>
1	<i>Cluster 5</i>	<i>impulsive, hearty, enthusiastic, resilient, optimistic, generous, vivacious</i>	<i>hearty, resilient, vivacious</i>
1	<i>Cluster 6</i>	<i>sentimental, idealistic, emotional</i>	<i>sentimental</i>
1	<i>Cluster 7</i>	<i>imaginary, hypothetical, ideal, subjective, conceptual, abstract, visionary</i>	<i>hypothetical, visionary</i>

1	Cluster 8	<i>sour, sensitive, inhibited, reclusive, high-strung, excitable, reserved, irritable</i>	<i>reserved, irritable</i>
1	Cluster 9	<i>open-hearted, frank, devout, honest, pious, religious, reverent</i>	<i>devout, honest</i>
1	Cluster 10	<i>constructive, sophisticated, practical</i>	<i>constructive</i>
1	Cluster 11	<i>energetic, in shape, influential, original, strong, alcoholic, courageous, fresh</i>	<i>strong, courageous</i>
1	Cluster 12	<i>selfish, arrogant, self-satisfied, self-confident, opinionated, imitative, boastful, egotistical, autocratic, affected, conceited</i>	<i>conceited, self-satisfied</i>
1	Cluster 13	<i>mindful, conjectural, academic, doubtful, self-conscious, thoughtful, cautious, speculative</i>	<i>thoughtful</i>
1	Cluster 14	<i>serious, ascetic, ambitious, industrious, hard, inflexible, dignified, exact, austere, formal, conventional, eloquent, elevated, severe, enterprising</i>	<i>ambitious, severe, dignified</i>
1	Cluster 15	<i>afraid, worried, frightened, panicky, coward, timid, anxious</i>	<i>afraid, anxious</i>
1	Cluster 16	<i>frugal, thrifty, sparing, provident, economical</i>	<i>frugal</i>
2	Cluster 17	<i>fancy</i>	<i>fancy</i>
3	Cluster 18	<i>fastidious</i>	<i>fastidious</i>
4	Cluster 19	<i>flattering</i>	<i>flattering</i>
5	Cluster 20	<i>preventive</i>	<i>preventive</i>
6	Cluster 21	<i>corporeal, corporeal, bodily, physical, animal</i>	<i>physical</i>
6	Cluster 22	<i>leading, interested, visible, curious, perceivable, prominent</i>	<i>prominent, interested</i>
7	Cluster 23	<i>acquisitive</i>	<i>acquisitive</i>
8	Cluster 24	<i>amorous</i>	<i>amorous</i>

9	Cluster 25	<i>pugnacious, argumentative</i>	<i>argumentative</i>
10	Cluster 26	<i>speaking, assertive, vocal, unwritten</i>	<i>assertive</i>
11	Cluster 27	<i>athletic</i>	<i>athletic</i>
12	Cluster 28	<i>closed</i>	<i>closed</i>
13	Cluster 29	<i>complaining, sarcastic, cynical</i>	<i>cynical</i>
14	Cluster 30	<i>tactful, painstaking, conscientious</i>	<i>conscientious</i>
15	Cluster 31	<i>contented</i>	<i>contented</i>
16	Cluster 32	<i>persevering, patient, controlled</i>	<i>patient</i>
17	Cluster 33	<i>musical, decorative</i>	<i>decorative, musical</i>
18	Cluster 34	<i>easygoing</i>	<i>easygoing</i>
19	Cluster 35	<i>eccentric</i>	<i>eccentric</i>
20	Cluster 36	<i>effeminate</i>	<i>effeminate</i>
21	Cluster 37	<i>evasive</i>	<i>evasive</i>
22	Cluster 38	<i>gluttonous</i>	<i>gluttonous</i>
23	Cluster 39	<i>governing</i>	<i>governing</i>
24	Cluster 40	<i>habituated</i>	<i>habituated</i>
25	Cluster 41	<i>mulish, headstrong</i>	<i>mulish</i>
26	Cluster 42	<i>hurried</i>	<i>hurried</i>
27	Cluster 43	<i>independent</i>	<i>independent</i>
28	Cluster 44	<i>influencing</i>	<i>influencing</i>
29	Cluster 45	<i>inhuman</i>	<i>inhuman</i>
30	Cluster 46	<i>introspective</i>	<i>introspective</i>
31	Cluster 47	<i>jealous</i>	<i>jealous</i>
32	Cluster 48	<i>laughing</i>	<i>laughing</i>
33	Cluster 49	<i>leader</i>	<i>leader</i>
34	Cluster 50	<i>verbal, lexical</i>	<i>verbal</i>
35	Cluster 51	<i>mischievous</i>	<i>mischievous</i>
36	Cluster 52	<i>mystical</i>	<i>mystical</i>
37	Cluster 53	<i>neurotic</i>	<i>neurotic</i>
38	Cluster 54	<i>observer</i>	<i>observer</i>
39	Cluster 55	<i>pedantic</i>	<i>pedantic</i>
40	Cluster 56	<i>perfidious</i>	<i>perfidious</i>
41	Cluster 57	<i>perfunctory</i>	<i>perfunctory</i>

42	<i>Cluster 58</i>	<i>pictorial</i>	<i>pictorial</i>
43	<i>Cluster 59</i>	<i>pitying</i>	<i>pitying</i>
44	<i>Cluster 60</i>	<i>plaintive</i>	<i>plaintive</i>
45	<i>Cluster 61</i>	<i>reliable, stable</i>	<i>reliable</i>
46	<i>Cluster 62</i>	<i>slanderosus</i>	<i>slanderosus</i>
47	<i>Cluster 63</i>	<i>slow</i>	<i>slow</i>
48	<i>Cluster 64</i>	<i>splenetic</i>	<i>splenetic</i>
49	<i>Cluster 65</i>	<i>sportive</i>	<i>sportive</i>
50	<i>Cluster 66</i>	<i>talkative</i>	<i>talkative</i>
51	<i>Cluster 67</i>	<i>temperate</i>	<i>temperate</i>
52	<i>Cluster 68</i>	<i>vindictive</i>	<i>vindictive</i>
53	<i>Cluster 69</i>	<i>visual</i>	<i>visual</i>
54	<i>Cluster 70</i>	<i>wandering</i>	<i>wandering</i>

* As células em vermelho representam components conexos do grafo compostos por vértices (traços) isolados.

O primeiro aspecto que de imediato se nota é que, com exceção dos vértices originalmente isolados, os *clusters* encontrados por esta metodologia foram todos compostos por mais de um elemento. Como mencionado anteriormente, isto é desejável, pois *clusters* com cardinalidade maior do que um pode gozar da propriedade de que é sempre possível se encontrar sentidos mais amplos para as palavras contidas em cada um deles (o que é desejável do ponto de vista de se organizar tais palavras).

Outra diferença em relação aos resultados obtidos com rede neurais SOM é que o número de *clusters* obtidos por MCL foi sensivelmente menor do que os encontrados na primeira metodologia (70 *clusters* contra 122 encontrados por SOM). Deste modo, tem-se que 19,37% das palavras que se encontram isoladas em *clusters* compostos só por elas correspondem somente aos traços que já estavam isolados dos demais no conjunto de traços selecionados.

Nenhuma palavra cujo vértice correspondente pertence a um componente conexo formado por mais de um vértice foi alocada em um *cluster* de cardinalidade unitária.

A seguir, se apresentará algumas análises dos *clusters* compostos por mais de uma palavra.

Cluster 1: composto por *intelligent, advisable, alert, logical, shrewd, witty, proficient, clever, skilled, wise, dexterous, professional, active, prepared, skilful, analytical, rational, observing* e *deft*. Está associado às ideias de ser esperto, inteligente e analítico. Portanto se escolheu as palavras *clever* e *analytical* para representar esses sentidos amplos deste agrupamento.

Cluster 2: Contêm os termos *discriminating, imaginative, multifarious, eclectic, many-sided, multiple*. Está associado às ideias gerais de ser eclético, multifacetado. Assim, se escolheu o termo *eclectic* para designar estas ideias gerais.

Cluster 3: Contém os termos *cosmopolitan, social, cordial, hospitable, cheerful, gregarious, poised, kind, aesthetic, coherent, kind-hearted, friendly, affable, affectionate, charming, polished, sociable, grateful, political, amiable, convivial, impartial, genial, cooperative, responsive, informative*. Está associado às ideias gerais de ser sociável e amigável. Assim, se escolheu os termos *sociable* e *amiable* para designarem estas ideias gerais.

Cluster 4: Formado por *programmed, mechanical, standardized, instinctive, intuitive*. Estão associadas às ideias gerais de ser instintivo e de se fazer as coisas mecanicamente. Assim, se escolheu o termo *instinctive* para designar estas ideias gerais.

Cluster 5: Composto por *impulsive, hearty, enthusiastic, resilient, optimistic, generous, vivacious*. Está associado às ideias de se ser raçudo, relente otimista e cheio de vida. Deste modo, se escolheu os termos *hearty, resilient* e *vivacious* para designarem estas ideias gerais.

Cluster 6: Composto por *sentimental, idealistic, emotional*. Está associado à ideia de se ser sentimental. Deste modo, se escolheu o termo *sentimental* para designar esta ideia geral.

Cluster 7: Formado por *imaginary, hypothetical, ideal, subjective, conceptual, abstract, visionary*. Está associado às ideias de se ser hipotético e visionário. Deste modo, se escolheu os termos *hypothetical* e *visionary* para designarem estas ideias gerais.

Cluster 8: Formado por *sour, sensitive, inhibited, reclusive, high-strung, excitable, reserved, irritable*. Está associado às ideias de se ser reservado e irritável. Deste modo, se escolheu os termos *reserved* e *irritable* para designarem estas ideias gerais.

Cluster 9: Formado por *open-hearted, frank, devout, honest, pious, religious, reverent*. Está associado às ideias de se ser honesto, franco e devoto. Deste modo, se escolheu os termos *devout* e *honest* para designarem estas ideias gerais.

Cluster 10: Formado por *constructive, sophisticated, practical*. Está associado às ideias gerais de ser construtivo e prático. Assim, se escolheu o termo *constructive* para designar estas ideias gerais.

Cluster 11: Compreendendo os termos *energetic, in shape, influential, original, strong, alcoholic, courageous, fresh*. Este agrupamento está relacionado às ideias gerais de

ser forte, energético e corajoso. Assim, se escolheu os termos *strong* e *courageous* para designarem estas ideias gerais.

Cluster 12: Compreendendo os termos *selfish, arrogant, self-satisfied, self-confident, opinionated, imitative, boastful, egotistical, autocratic, affected, conceited*. Este agrupamento está relacionado às ideias gerais de ser vaidoso, convencido e de se estar satisfeito consigo mesmo. Assim, se escolheu os termos *conceited* e *self-satisfied* para designarem estas ideias gerais.

Cluster 13: Compreendendo os termos *mindful, conjectural, academic, doubtful, self-conscious, thoughtful, cautious, speculative*. Este agrupamento está relacionado às ideias gerais de ser pensativo, conjectural e especulativo. Assim, se escolheu o termo *thoughtful* para designar estas ideias gerais.

Cluster 14: Compreendendo os termos *serious, ascetic, ambitious, industrious, hard, inflexible, dignified, exact, austere, formal, conventional, eloquent, elevated, severe e enterprising*. Este agrupamento está relacionado às ideias gerais de ser ambicioso, severo e dignificado. Assim, se escolheu os termos *ambitious, severe* e *dignified* para designarem estas ideias gerais.

Cluster 15: Compreende os traços *afraid, worried, frightened, panicky, coward, timid* e *anxious*. Este agrupamento está relacionado às ideias gerais de ser medroso, preocupado e ansioso. Assim, designou-se *afraid* e *anxious* para designarem estas ideias.

Cluster 16: Compreende os traços *frugal, thrifty, sparing, provident* e *economical*. Este conjunto está relacionado às ideias gerais de ser frugal, providente e econômico. Assim, designou-se *frugal* para designar estas ideias.

Cluster 21: Compreende os traços *corporais, corporeal, bodily, physical* e *animal*. Está relacionado às ideias gerais de ser animal, corpóreo e físico. Designou-se *physical* para designar estas ideias.

Cluster 22: Compreende os traços *leading, interested, visible, curious, perceivable* e *prominent*. Está relacionado às ideias gerais de ser interessado, curioso, proeminente e de se estar na vanguarda. Designaram-se os termos *prominent* e *interested* para designar estas ideias.

Cluster 25: Compreende os traços *pugnacious* e *argumentative*. Está relacionado às ideias gerais de ser pugnaz, brigão e argumentativo. Designou-se o termo *argumentative* para designar estas ideias.

Cluster 26: Engloba os traços *speaking, assertive, vocal* e *unwritten*. Está relacionado às ideias gerais de ser falante, assertivo e vocal. Designou-se o termo *assertivo* para designar estas ideias.

Cluster 29: Englobando os traços *complaining, sarcastic* e *cynical*. Está relacionado às ideias gerais de ser ranzinza, sarcástico e cínico. Designou-se o termo *cínico* para designar estas ideias.

Cluster 30: Englobando os traços *tactful, painstaking* e *conscientious*. Está relacionado às ideias gerais de se ter tato e ser consciente. Designou-se o termo *conscientious* para designar estas ideias.

Cluster 32: Contendo os traços *persevering, patient* e *controlled*. Está relacionado às ideias gerais de se ser paciente, perseverante e controlado. Designou-se o termo *patient* para designar estas ideias.

Cluster 33: Contendo os traços *musical* e *decorative*. Está relacionado às ideias gerais de se ser musical e paciente. Designaram-se os mesmos termos para representarem as ideias desse *cluster*.

Cluster 41: Contendo os traços *mulish* e *headstrong*. Está relacionado às ideias gerais de se ser teimoso e paciente. Designou-se o termo *mulish* para representar as ideias desse *cluster*.

Cluster 50: Contendo os traços *verbal* e *lexical*. Está relacionado às ideias gerais de se ser verbal e léxico (de gostar de se expressar e de verbalizar suas experiências). Designou-se o termo *verbal* para representar as ideias desse *cluster*.

Cluster 61: Contendo os traços *reliable* e *stable*. Está relacionado às ideias gerais de se ser confiável e estável. Designou-se o termo *reliable* para representar as ideias desse *cluster*.

Os demais *clusters* são formados por um único elemento.

Como também pode ser verificado, boa parte dos *clusters* encontrados apresentaram alto grau de conectividade entre suas palavras-componentes e menor conectividade inter *cluster*. Em se tratando de agrupamentos obtidos pelo algoritmo MCL, este resultado era esperado, haja vista que o mesmo opera de forma a privilegiar regiões de maior conectividade entre vértices para definir *clusters*.

As figuras 1 a 8 no Apêndice apresentam os *clusters* identificados nas diversas componentes conexas do grafo, compostos por mais de um traço, segundo a metodologia MCL.

Tal como realizado nas análises dos resultados obtidos com redes SOM, também se procedeu à investigação dos percentuais de ligações *intra* e *interclusters* para os grupos identificados pelo algoritmo MCL. As tabelas 30 e 31 abaixo apresentam estes resultados.

Tabela 30: Percentuais de ligações *intra* e *interclusters* verificados para os resultados obtidos com o algoritmo MCL na componente conexa 1 do grafo.

Resultados MCL								
Compo. Conexa do grafo	Cluster #	Rótulos dos clusters	# traços	Total de ligações dos traços do cluster	# de ligações intra cluster	# de ligações inter clusters	% de ligações intra clusters	% de ligações inter clusters
1	1	<i>clever, analytical</i>	19	34	30	4	88,24%	11,76%
1	2	<i>eclectic</i>	6	10	6	4	60,00%	40,00%
1	3	<i>sociable, amiable</i>	26	44	38	6	86,36%	13,64%
1	4	<i>instinctive</i>	5	5	4	1	80,00%	20,00%
1	5	<i>hearty, resilient, vivacious</i>	7	20	7	13	35,00%	65,00%
1	6	<i>sentimental</i>	3	5	2	3	40,00%	60,00%
1	7	<i>hypothetical, visionary</i>	7	11	9	2	81,82%	18,18%
1	8	<i>reserved, irritable</i>	8	10	9	1	90,00%	10,00%
1	9	<i>devout, honest</i>	7	8	7	1	87,50%	12,50%
1	10	<i>constructive</i>	3	5	2	3	40,00%	60,00%
1	11	<i>strong, courageous</i>	8	9	7	2	77,78%	22,22%
1	12	<i>conceited, self-satisfied</i>	11	15	14	1	93,33%	6,67%
1	13	<i>thoughtful</i>	8	9	8	1	88,89%	11,11%
1	14	<i>ambitious, severe, dignified</i>	15	21	17	4	80,95%	19,05%
1	15	<i>afraid, anxious</i>	7	8	7	1	87,50%	12,50%
1	16	<i>frugal</i>	5	10	9	1	90,00%	10,00%
Valores médios na comp. 1:			9,0625	14	11	3	75,46%	24,54%

Nota: na tabela acima não foram considerados os traços pertencentes a componentes conexas originalmente compostas somente por um único termo.

Tabela 31: Percentuais de ligações *intra* e *interclusters* verificados para os resultados obtidos com o algoritmo MCL nas demais componentes conexas do grafo.

Resultados MCL								
Compo. Conexo do grafo	Cluster #	Rótulos dos clusters	# traços	Total de ligações dos traços do cluster	# de ligações <i>intra cluster</i>	# de ligações <i>inter clusters</i>	% de ligações <i>intra clusters</i>	% de ligações <i>inter clusters</i>
34	50	<i>verbal</i>	2	1	1	0	100,00%	0,00%
25	41	<i>mulish</i>	2	1	1	0	100,00%	0,00%
45	61	<i>reliable</i>	2	1	1	0	100,00%	0,00%
17	33	<i>decorative, musical</i>	2	1	1	0	100,00%	0,00%
16	32	<i>patient</i>	2	2	2	0	100,00%	0,00%
14	30	<i>conscientious</i>	3	2	2	0	100,00%	0,00%
13	29	<i>cynical</i>	3	2	2	0	100,00%	0,00%
10	26	<i>assertive</i>	4	3	3	0	100,00%	0,00%
9	25	<i>argumentative</i>	2	1	1	0	100,00%	0,00%
6	22	<i>prominent, interested</i>	6	6	5	1	83,33%	16,67%
6	21	<i>physical</i>	5	8	7	1	87,50%	12,50%

Nota: na tabela acima não foram considerados os traços pertencentes a componentes conexas originalmente compostas somente por um único termo.

Como se pode observar nas tabelas 30 e 31 acima, nos *clusters* identificados por MCL predominaram ligações *intra clusters* sobre as ligações *inter clusters*, obtendo-se, neste sentido, um resultado praticamente inverso ao verificado nos *clusters* encontrados pela rede SOM.

Este resultado já era esperado, em se tratando do algoritmo MCL, uma vez que para o mesmo, regiões de maior concentração de ligações entre vértices são candidatas naturais a formarem *clusters*.

Deste modo, com os parâmetros escolhidos para ambos os algoritmos de clusterização, caso se desejasse impor como restrição aos clusters encontrados por ambas as metodologias a condição de que a maioria das ligações entre vértices do grafo sejam *intra clusters*, se recomendaria a adoção dos *clusters* obtidos por MCL e haveria necessidade de se mudar os parâmetros da rede SOM (e/ou se proceder a outra estratégia de codificação dos vértices do grafo; por exemplo, se procedendo a uma codificação independente em cada componente conexo do grafo contendo mais de um elemento).

7.3 COMPARAÇÃO DOS RESULTADOS OBTIDOS ENTRE AS DIFERENTES METODOLOGIAS

Os resultados obtidos com o algoritmo MCL e com a rede neural SOM são justapostos na tabela a seguir.

Tabela 32: Comparação entre os *clusters* obtidos com redes SOM e MCL.

Resultados com MCL. Inflation = 1,85				Resultados com SOM						
GRUPO CONEXO	Clusters	Palavras	Conceitos/idéias gerais associados a cada cluster - MCL	GRUPO CONEXO	Clusters	Palavras				
1	Cluster 1	<i>intelligent, advisable, alert, logical, shrewd, witty, proficient, clever, skilled, wise, dexterous, professional, active, prepared, skilful, analytical, rational, observing, deft</i>	<i>clever, analytical</i>	1	Cluster 3	<i>clever, dexterous, witty</i>				
				1	Cluster 4	<i>skilful, deft, skilled, prepared, proficient, professional</i>				
				1	Cluster 5	<i>alert, fresh, original</i>				
				1	Cluster 17	<i>rational, analytical, advisable</i>				
				1	Cluster 18	<i>intelligent, logical, wise</i>				
				1	Cluster 19	<i>shrewd, hard</i>				
				1	Cluster 31	<i>active</i>				
				1	Cluster 60	<i>observing</i>				
				1	Cluster 2	<i>discriminating, imaginative, multifarious, eclectic, many-sided, multiple</i>	<i>eclectic</i>	1	Cluster 7	<i>discriminating, imaginative, aesthetic, polished, informative, hospitable, poised</i>
								1	Cluster 21	<i>eclectic, multifarious</i>
1	Cluster 35	<i>many-sided, multiple</i>								

1	Cluster 3	<i>cosmopolitan, social, cordial, hospitable, gregarious, poised, kind, aesthetic, coherent, kind-hearted, friendly, affable, affectionate, charming, polished, sociable, grateful, political, amiable, convivial, impartial, genial, cooperative, responsive, informative</i>	<i>sociable, amiable</i>	1	Cluster 8	<i>social, sociable, gregarious, convivial, genial, cordial, friendly, amiable, affable, political, cooperative, coherent, cosmopolitan, impartial, cheerful, resilient, optimistic, affectionate, kind, responsive, kind-hearted, charming, grateful</i>
1	Cluster 4	<i>programmed, mechanical, standardized, instinctive, intuitive</i>	<i>instinctive</i>	1	Cluster 59	<i>mechanical, programmed</i>
				1	Cluster 70	<i>standardized</i>
				1	Cluster 71	<i>instinctive, intuitive</i>
1	Cluster 5	<i>impulsive, hearty, enthusiastic, resilient, optimistic, generous, vivacious, cheerful,</i>	<i>hearty, resilient, vivacious</i>	1	Cluster 82	<i>impulsive</i>
				1	Cluster 16	<i>hearty, generous</i>
				1	Cluster 105	<i>enthusiastic</i>
				1	Cluster 94	<i>vivacious</i>
1	Cluster 6	<i>sentimental, idealistic, emotional</i>	<i>sentimental</i>	1	Cluster 83	<i>emotional</i>
				1	Cluster 84	<i>idealistic</i>
				1	Cluster 95	<i>sentimental</i>
1	Cluster 7	<i>imaginary, hypothetical, ideal, subjective, conceptual, abstract, visionary</i>	<i>hypothetical, visionary</i>	1	Cluster 46	<i>visionary, ideal, hypothetical</i>
				1	Cluster 85	<i>imaginary, subjective</i>
				1	Cluster 86	<i>abstract</i>
				1	Cluster 96	<i>conceptual</i>
1	Cluster 8	<i>sour, sensitive, inhibited, reclusive, high-strung, excitable, reserved, irritable</i>	<i>reserved, irritable</i>	1	Cluster 47	<i>reclusive, reserved, inhibited</i>
				1	Cluster 104	<i>irritable</i>
				1	Cluster 106	<i>excitable</i>
				1	Cluster 114	<i>sour</i>
				1	Cluster 115	<i>sensitive</i>
				1	Cluster 116	<i>high-strung</i>
1	Cluster 9	<i>open-hearted, frank, devout, honest, pious, religious, reverent</i>	<i>devout, honest</i>	1	Cluster 1	<i>frank, open-hearted</i>

				1	Cluster 2	<i>devout, honest, religious, pious, reverent</i>
1	Cluster 10	<i>constructive, sophisticated, practical</i>	<i>constructive</i>	1	Cluster 20	<i>practical, sophisticated, constructive</i>
1	Cluster 11	<i>energetic, in shape, influential, original, strong, alcoholic, courageous, fresh</i>	<i>strong, courageous</i>	1	Cluster 29	<i>strong, alcoholic, energetic, courageous, influential</i>
				1	Cluster 30	<i>in shape</i>
1	Cluster 12	<i>selfish, arrogant, self-satisfied, self-confident, opinionated, boastful, egotistical, autocratic, affected, conceited</i>	<i>conceited, self-satisfied</i>	1	Cluster 75	<i>conceited, arrogant, egotistical, imitative, autocratic, opinionated, selfish</i>
				1	Cluster 62	<i>self-satisfied, self-confident</i>
				1	Cluster 74	<i>boastful</i>
				1	Cluster 87	<i>affected</i>
1	Cluster 13	<i>mindful, conjectural, academic, doubtful, self-conscious, thoughtful, cautious, speculative</i>	<i>thoughtful</i>	1	Cluster 45	<i>thoughtful, mindful, cautious</i>
				1	Cluster 61	<i>doubtful, self-conscious</i>
				1	Cluster 72	<i>speculative, conjectural</i>
				1	Cluster 73	<i>academic</i>
1	Cluster 14	<i>serious, ascetic, ambitious, industrious, hard, inflexible, dignified, exact, austere, formal, conventional, eloquent, elevated, severe, enterprising</i>	<i>ambitious, severe, dignified</i>	1	Cluster 33	<i>severe, exact, ambitious, ascetic, serious, enterprising, industrious</i>
				1	Cluster 32	<i>inflexible</i>
				1	Cluster 34	<i>austere, formal</i>
				1	Cluster 49	<i>conventional</i>
				1	Cluster 48	<i>dignified, elevated, eloquent</i>
1	Cluster 15	<i>afraid, worried, frightened, panicky, coward, timid, anxious</i>	<i>afraid, anxious</i>	1	Cluster 97	<i>worried, frightened</i>
				1	Cluster 107	<i>panicky</i>
				1	Cluster 108	<i>timid</i>
				1	Cluster 117	<i>anxious</i>
				1	Cluster 118	<i>afraid</i>
				1	Cluster 119	<i>coward</i>

1	Cluster 16	<i>frugal, thrifty, sparing, provident, economical</i>	<i>frugal</i>	1	Cluster 6	<i>economical, frugal, provident, thrifty, sparing</i>
2	Cluster 17	<i>fancy</i>	<i>fancy</i>	2	Cluster 12	<i>fancy</i>
3	Cluster 18	<i>fastidious</i>	<i>fastidious</i>	3	Cluster 90	<i>fastidious</i>
4	Cluster 19	<i>flattering</i>	<i>flattering</i>	4	Cluster 76	<i>flattering</i>
5	Cluster 20	<i>preventive</i>	<i>preventive</i>	5	Cluster 24	<i>preventive</i>
6	Cluster 21	<i>corporeal, corporeal, bodily, physical, animal</i>	<i>physical</i>	6	Cluster 92	<i>physical</i>
				6	Cluster 112	<i>corporeal</i>
				6	Cluster 121	<i>corporeal</i>
				6	Cluster 91	<i>bodily</i>
				6	Cluster 79	<i>animal</i>
6	Cluster 22	<i>leading, interested, visible, curious, perceivable, prominent</i>	<i>prominent, interested</i>	6	Cluster 81	<i>prominent</i>
				6	Cluster 93	<i>visible</i>
				6	Cluster 102	<i>curious</i>
				6	Cluster 103	<i>leading</i>
				6	Cluster 113	<i>interested</i>
				6	Cluster 122	<i>perceivable</i>
7	Cluster 23	<i>acquisitive</i>	<i>acquisitive</i>	7	Cluster 38	<i>acquisitive</i>
8	Cluster 24	<i>amorous</i>	<i>amorous</i>	8	Cluster 40	<i>amorous</i>
9	Cluster 25	<i>pugnacious, argumentative</i>	<i>argumentative</i>	9	Cluster 23	<i>argumentative, pugnacious</i>
10	Cluster 26	<i>speaking, assertive, vocal, unwritten</i>	<i>assertive</i>	10	Cluster 14	<i>assertive, vocal, speaking, unwritten</i>
11	Cluster 27	<i>athletic</i>	<i>athletic</i>	11	Cluster 64	<i>athletic</i>
12	Cluster 28	<i>closed</i>	<i>closed</i>	12	Cluster 65	<i>closed</i>
13	Cluster 29	<i>complaining, sarcastic, cynical</i>	<i>cynical</i>	13	Cluster 28	<i>complaining, cynical, sarcastic</i>
14	Cluster 30	<i>tactful, painstaking, conscientious</i>	<i>conscientious</i>	14	Cluster 15	<i>conscientious, painstaking, tactful</i>
15	Cluster 31	<i>contented</i>	<i>contented</i>	15	Cluster 66	<i>contented</i>
16	Cluster 32	<i>persevering, patient, controlled</i>	<i>patient</i>	16	Cluster 44	<i>controlled, patient</i>
				16	Cluster 58	<i>persevering</i>
17	Cluster 33	<i>musical, decorative</i>	<i>decorative, musical</i>	17	Cluster 120	<i>musical</i>
				17	Cluster 109	<i>decorative</i>
18	Cluster 34	<i>easygoing</i>	<i>easygoing</i>	18	Cluster 36	<i>easygoing</i>
19	Cluster 35	<i>eccentric</i>	<i>eccentric</i>	19	Cluster 78	<i>eccentric</i>

20	Cluster 36	<i>effeminate</i>	<i>effeminate</i>	20	Cluster 88	<i>effeminate</i>
21	Cluster 37	<i>evasive</i>	<i>evasive</i>	21	Cluster 68	<i>evasive</i>
22	Cluster 38	<i>gluttonous</i>	<i>gluttonous</i>	22	Cluster 110	<i>gluttonous</i>
23	Cluster 39	<i>governing</i>	<i>governing</i>	23	Cluster 26	<i>governing</i>
24	Cluster 40	<i>habituated</i>	<i>habituated</i>	24	Cluster 39	<i>habituated</i>
25	Cluster 41	<i>mulish, headstrong</i>	<i>mulish</i>	25	Cluster 99	<i>headstrong, mulish</i>
26	Cluster 42	<i>hurried</i>	<i>hurried</i>	26	Cluster 67	<i>hurried</i>
27	Cluster 43	<i>independent</i>	<i>independent</i>	27	Cluster 63	<i>independent</i>
28	Cluster 44	<i>influencing</i>	<i>influencing</i>	28	Cluster 13	<i>influencing</i>
29	Cluster 45	<i>inhuman</i>	<i>inhuman</i>	29	Cluster 50	<i>inhuman</i>
30	Cluster 46	<i>introspective</i>	<i>introspective</i>	30	Cluster 37	<i>introspective</i>
31	Cluster 47	<i>jealous</i>	<i>jealous</i>	31	Cluster 98	<i>jealous</i>
32	Cluster 48	<i>laughing</i>	<i>laughing</i>	32	Cluster 51	<i>laughing</i>
33	Cluster 49	<i>leader</i>	<i>leader</i>	33	Cluster 69	<i>leader</i>
34	Cluster 50	<i>verbal, lexical</i>	<i>verbal</i>	34	Cluster 80	<i>lexical, verbal</i>
35	Cluster 51	<i>mischievous</i>	<i>mischievous</i>	35	Cluster 22	<i>mischievous</i>
36	Cluster 52	<i>mystical</i>	<i>mystical</i>	36	Cluster 100	<i>mystical</i>
37	Cluster 53	<i>neurotic</i>	<i>neurotic</i>	37	Cluster 101	<i>neurotic</i>
38	Cluster 54	<i>observer</i>	<i>observer</i>	38	Cluster 25	<i>observer</i>
39	Cluster 55	<i>pedantic</i>	<i>pedantic</i>	39	Cluster 89	<i>pedantic</i>
40	Cluster 56	<i>perfidious</i>	<i>perfidious</i>	40	Cluster 10	<i>perfidious</i>
41	Cluster 57	<i>perfunctory</i>	<i>perfunctory</i>	41	Cluster 42	<i>perfunctory</i>
42	Cluster 58	<i>pictorial</i>	<i>pictorial</i>	42	Cluster 9	<i>pictorial</i>
43	Cluster 59	<i>pitying</i>	<i>pitying</i>	43	Cluster 41	<i>pitying</i>
44	Cluster 60	<i>plaintive</i>	<i>plaintive</i>	44	Cluster 111	<i>plaintive</i>
45	Cluster 61	<i>reliable, stable</i>	<i>reliable</i>	45	Cluster 43	<i>reliable, stable</i>
46	Cluster 62	<i>slandorous</i>	<i>slandorous</i>	46	Cluster 54	<i>slandorous</i>
47	Cluster 63	<i>slow</i>	<i>slow</i>	47	Cluster 53	<i>slow</i>
48	Cluster 64	<i>splenetic</i>	<i>splenetic</i>	48	Cluster 57	<i>splenetic</i>
49	Cluster 65	<i>sportive</i>	<i>sportive</i>	49	Cluster 77	<i>sportive</i>
50	Cluster 66	<i>talkative</i>	<i>talkative</i>	50	Cluster 56	<i>talkative</i>
51	Cluster 67	<i>temperate</i>	<i>temperate</i>	51	Cluster 52	<i>temperate</i>
52	Cluster 68	<i>vindictive</i>	<i>vindictive</i>	52	Cluster 11	<i>vindictive</i>
53	Cluster 69	<i>visual</i>	<i>visual</i>	53	Cluster 55	<i>visual</i>
54	Cluster 70	<i>wandering</i>	<i>wandering</i>	54	Cluster 27	<i>wandering</i>

* As células em vermelho na tabela acima representam os componentes conexos, só grafo formado por apenas um único vértice (traço) isolado.

As diferenças nos resultados encontrados pelas duas metodologias refletem em parte as diferenças de critérios utilizados por estes algoritmos na identificação de *clusters*: no MCL regiões com maior concentração de ligações entre as palavras (ou maior concentração de arestas no grafo correspondente) são candidatas mais prováveis na formação destes agrupamentos.

Com redes SOM, a definição final dos *clusters* irá depender de uma série de fatores, como a escolha da arquitetura do mapa topológico - incluindo o número de polígonos e como estarão dispostos neste mapa-, a dimensão do mapa, as definições de como as vizinhanças no entorno de cada nó no mapa serão definidas e de como serão atualizadas em cada iteração, a taxa de aprendizagem que a rede utilizará etc.).

Esta definição com redes SOM, no problema abordado, irá depender também de como os vetores que codificam cada vértice do grafo sendo clusterizado são definidos. Mais especificamente está-se referindo aqui às escolhas feitas para os elementos não nulos das componentes adicionais que são incluídas nos vetores para atribuição de distâncias explosivas entre vértices pertencentes a diferentes componentes conexos do grafo.

Finalmente, os cluster finais com as redes de Kohonen irão ser influenciados também pela medida de distância utilizada (novamente, nas aplicações apresentadas acima com redes SOM se utilizou a distância Euclidiana).

Contudo, acredita-se que se pode alcançar resultados de clusterização mais próximos entre as metodologias com outras escolhas dos parâmetros mencionados.

Independentemente das diferenças, ambas as metodologias de clusterização foram capazes de encontrar agrupamentos de traços concisos. Nestes agrupamentos a identificação dos conceitos/ideias gerais presentes em cada um deles foi um processo de fácil execução.

7.4 COMPARAÇÃO DOS RESULTADOS DE CLUSTERIZAÇÃO OBTIDOS COM O MODELO DE 12 FATORES DE PERSONALIDADE DE RAYMOND CATTELL (1945b)

Embora esta tese não tenha como objetivo a construção de um modelo de personalidade, se irá proceder aqui a uma comparação dos CTP's obtidos com as duas metodologias de clusterização com os 12 fatores do modelo de personalidade identificado por Raymond Cattell (1945 e 1947).

Antes de se proceder a essas comparações, deve-se lembrar de que o modelo de 12 fatores da personalidade deste autor foi construído com a seguinte estrutura:

- O autor partiu de um conjunto de 171 traços selecionados a partir de um amplo conjunto de termos agrupados como sinônimos.

- A partir da correlação entre os *ratings* dados em respostas a questionários de avaliação de indivíduos contendo esses 171 traços construíram *clusters*, a partir dos quais conseguiu reduzir ainda mais esse conjunto de traços para um conjunto final de 105, agrupados em 35 *clusters*, que passaram a representar as variáveis a serem utilizadas no experimento de análise fatorial a partir das correlações entre *ratings* de respostas a questionários contendo esses 105 traços aplicados em outra população de teste.

Com base nestas correlações o autor conseguiu identificar 12 fatores (cada um desses fatores representando um subconjunto de variáveis mais fortemente correlacionadas entre si).

Os *clusters* obtidos com a aplicação dos dois algoritmos de clusterização (redes SOM e MCL) no conjunto dos 222 traços de personalidade presentes no *thesaurus* adotado representam os agrupamentos desses termos de acordo com a linguagem natural.

A comparação dos fatores do supracitado modelo de Cattell com estes *clusters* visaria identificar como a linguagem natural conseguiu retratar (tal como descrita no *thesaurus* utilizado), com maior ou menor fidelidade, os fatores de personalidade do modelo (que por sua vez representam variáveis - ou *clusters* de traços identificados pelo autor, segundo as respostas em questionários - mais correlacionadas entre si no que diz respeito aos comportamentos tal como avaliados nesses questionários).

Contudo, deve-se notar de imediato que esta comparação deveria ser feita, a princípio, entre o citado modelo de personalidade e os resultados de clusterização obtidos a partir do mesmo conjunto dos 105 traços por ele utilizado.

Como se viu anteriormente os experimentos de clusterização no *thesaurus* aqui efetuados envolveram um conjunto maior de traços (222 traços de personalidade); parte deles exatamente idênticos a alguns dos 171 termos inicialmente propostos por Cattell (1943a, 1943b) e parte candidatos a substitutos de alguns dos termos e ausentes no *thesaurus*.

Contudo, analisando-se a rede de ligações entre estes 222 termos (adotados nestes experimentos de clusterização) no TTT, se verificou que os traços que correspondem às palavras não consideradas por Cattell para inclusão no modelo de 12 fatores (ou que são

candidatas às substitutas dessas palavras que se encontravam ausentes no banco de palavras usado), no *thesaurus*, se constituem, em sua grande maioria, a vértices originalmente isolados, ou que possuem uma única ligação com outro termo; de forma que não é esperado que a remoção destes termos do conjunto inicial de 222 traços afete substancialmente os resultados de clusterização obtidos com as duas metodologias empregadas.

Assim, não se espera que a remoção de parte dos termos do conjunto inicial de 222 traços selecionados para as aplicações de clusterização no *thesaurus* (visando-se chegar a um conjunto menor, que representasse os 105 traços finais utilizados no modelo 12PF) afete substancialmente os resultados de clusterização aqui obtidos com todo o conjunto dos 222 traços; de modo que a comparação aqui efetuada entre os clusters obtidos a partir desse conjunto maior de traços e os 12 fatores do modelo ainda é útil para as análises que se seguirão.

Não obstante, recomenda-se como estudo futuro, a comparação dos fatores desse modelo com os *clusters* obtidos com um conjunto de cardinalidade similar à do conjunto de traços finais selecionados por Cattell (para a identificação dos mencionados 12 fatores) no *thesaurus* aqui utilizado.

Outro ponto que merece destaque é que, nas comparações entre os resultados de clusterização encontrados com ambas as metodologias empregadas e o modelo 12 PF foram necessárias à consideração das distâncias entre *clusters* (encontrados segundo cada uma das metodologias utilizadas), em uma mesma componente conexa do grafo associado à rede de ligações entre os traços considerados. Para tanto, nessas situações se utilizou como medida de distância entre *clusters* o critério correspondente ao de Ligação Mínima (ou *Single Linkage*), que implica que a distância entre dois *clusters* X e Y seja definida por:

$$\min_{\{x \in X, y \in Y\}} d(x, y)$$

Uma vez feitas estas considerações, apresenta-se nas tabelas 33 e 34 abaixo as relações entre os fatores do mencionado modelo, as variáveis pertencentes a cada um deles e os *clusters* encontrados por meio da aplicação dos algoritmos de clusterização no TTT.

Tabela 33: Comparação dos fatores do modelo 12 PF de Cattell (1945b), com os clusters obtidos com o algoritmo MCL.

Fatores do modelo 12PF (CATTEL 1945) contendo as 35 variáveis	Descrição dos polos do fator		Traço correspondente na seleção inicial de 171 (CATTELL, 1943b)	Traço correspondente e utilizado neste trabalho	Polos dos traços pertencentes ao modelo 12PF		Número da Variável (Cluster) contendo o traço (Catell - 1945)	Polos da Variável contendo o traço		Cluster MCL
D/E	<i>Hypersensitive, infantile, sthenic emotionality/Dominance (Hypomania)</i>	<i>Phlegmatic frustration tolerance/Submissiveness</i>	<i>boastful</i>	<i>boastful</i>	<i>Boastful</i>	<i>Modest</i>	1	<i>Self Assertive</i>	<i>Self Submissive</i>	12
D/E	<i>Hypersensitive, infantile, sthenic emotionality/Dominance (Hypomania)</i>	<i>Phlegmatic frustration tolerance/Submissiveness</i>	<i>conceited</i>	<i>conceited</i>	<i>Conceited</i>	<i>Self critical, dissatisfied</i>	1	<i>Self Assertive</i>	<i>Self Submissive</i>	12
D/E	<i>Hypersensitive, infantile, sthenic emotionality/Dominance (Hypomania)</i>	<i>Phlegmatic frustration tolerance/Submissiveness</i>	<i>assertive</i>	<i>assertive</i>	<i>Assertive</i>	<i>Submissive</i>	1	<i>Self Assertive</i>	<i>Self Submissive</i>	26
B/J	<i>Intelligence, General Mental Capacity/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Neurasthenia</i>	<i>Abilities: intelligence</i>	<i>intelligent</i>	<i>Intelligent</i>	<i>Stupid</i>	2	<i>Intelligent, analytical</i>	<i>Unimaginative, stupid</i>	1
B/J	<i>Intelligence, General Mental Capacity/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Neurasthenia</i>	<i>clever</i>	<i>clever</i>	<i>Clever</i>	-	2	<i>Intelligent, analytical</i>	<i>Unimaginative, stupid</i>	1
B/J	<i>Intelligence, General Mental Capacity/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Neurasthenia</i>	<i>clear thinking</i>	<i>coherent</i>	<i>Clear Thinking</i>	<i>Incoherent, confused</i>	2	<i>Intelligent, analytical</i>	<i>Unimaginative, stupid</i>	3
B/G/V/J	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Rigid, tough, poise/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality/Neurasthenia</i>	<i>independent</i>	<i>independent</i>	<i>Independent</i>	<i>Emotionally dependent</i>	3	<i>Wise, mature, polished</i>	<i>Dependent, silly, incoherent</i>	43

B/G/I/J	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Rigid, tough, poise/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality/Neurasthenia</i>	<i>reliable</i>	<i>reliable</i>	<i>Reliable</i>	<i>Undependable</i>	3	<i>Wise, mature, polished</i>	<i>Dependent, silly, incoherent</i>	61
B/G/I/J	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Rigid, tough, poise/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality/Neurasthenia</i>	<i>Mature (emotional)- (Não incluída(Não incluída na clusterização))</i>	-	<i>Mature</i>	<i>Emotionally immature, irresponsible</i>	3	<i>Wise, mature, polished</i>	<i>Dependent, silly, incoherent</i>	-
B/C/G	<i>Mental Defect/Demoralized/Dependent, silly, incoherent</i>	<i>Intelligence, General Mental Capacity/Realistic, facing life/Wise, mature, polished</i>	<i>impulsive temperamentally</i>	<i>impulsive</i>	<i>Impulsive</i>	<i>Deliberate</i>	4	<i>Changeable, frivolous</i>	<i>Thoughtful, stoic, reserved</i>	5
B/C/G	<i>Mental Defect/Demoralized/Dependent, silly, incoherent</i>	<i>Intelligence, General Mental Capacity/Realistic, facing life/Wise, mature, polished</i>	<i>thoughtful</i>	<i>thoughtful</i>	<i>Unreflective</i>	<i>Thoughtful</i>	4	<i>Changeable, frivolous</i>	<i>Thoughtful, stoic, reserved</i>	13
B/C/G	<i>Mental Defect/Demoralized/Dependent, silly, incoherent</i>	<i>Intelligence, General Mental Capacity/Realistic, facing life/Wise, mature, polished</i>	<i>austere</i>	<i>austere</i>	<i>Profligate</i>	<i>Austere</i>	4	<i>Changeable, frivolous</i>	<i>Thoughtful, stoic, reserved</i>	14
D/I	<i>Hypersensitive, infantile, sthenic emotionality/Sensitive, imaginative, anxious emotionality</i>	<i>Phlegmatic frustration tolerance/Rigid, tough, poise</i>	<i>hypochondriacal (Catell), hypochondriac</i>	<i>anxious</i>	<i>Hypochondriacal</i>	-	5	<i>Neurotic</i>	<i>Not generally neurotic</i>	15
D/I	<i>Hypersensitive, infantile, sthenic emotionality/Sensitive, imaginative, anxious emotionality</i>	<i>Phlegmatic frustration tolerance/Rigid, tough, poise</i>	<i>hypochondriacal (Catell), neurotic, hypochondriac</i>	<i>neurotic</i>	<i>Nervous, specific neurotic symptoms</i>	-	5	<i>Neurotic</i>	<i>Not generally neurotic</i>	53

D/I	<i>Hypersensitive, infantile, sthenic emotionality/Sensitive, imaginative, anxious emotionality</i>	<i>Phlegmatic frustration tolerance/Rigid, tough, poise</i>	<i>self-deceiving (Não incluída(Não incluída na clusterização)</i>	-	<i>Self-deceiving</i>	<i>Realistic</i>	5	<i>Neurotic</i>	<i>Not generally neurotic</i>	-
H/V/L	<i>Obstructive, withdrawn, schizothymia/Rigid, tough, poise/Paranoia</i>	<i>Charitable, adventurous, cyclothymia/Sensitive, imaginative, anxious emotionality/Surgent cyclothymia</i>	<i>grateful</i>	<i>grateful</i>	<i>Thankless</i>	<i>Grateful</i>	6	<i>Hard, cynical</i>	<i>Kindly, gentle, idealistic</i>	3
H/V/L	<i>Obstructive, withdrawn, schizothymia/Rigid, tough, poise/Paranoia</i>	<i>Charitable, adventurous, cyclothymia/Sensitive, imaginative, anxious emotionality/Surgent cyclothymia</i>	<i>friendly</i>	<i>friendly</i>	<i>Hostile</i>	<i>Friendly, understanding</i>	6	<i>Hard, cynical</i>	<i>Kindly, gentle, idealistic</i>	3
H/V/L	<i>Obstructive, withdrawn, schizothymia/Rigid, tough, poise/Paranoia</i>	<i>Charitable, adventurous, cyclothymia/Sensitive, imaginative, anxious emotionality/Surgent cyclothymia</i>	<i>hard</i>	<i>hard</i>	<i>Hardhearted</i>	<i>Softhearted</i>	6	<i>Hard, cynical</i>	<i>Kindly, gentle, idealistic</i>	14
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>extra-punitive</i>	<i>severe</i>	<i>Extra-punitive (blaming mistakes on others)</i>	-	7	<i>Wilful, egotistic, predatory</i>	<i>Mild, self-effacing, tolerant</i>	14
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>exhibitionistic</i>	<i>perceivable</i>	<i>Exhibitionist</i>	<i>Self-effacing</i>	7	<i>Wilful, egotistic, predatory</i>	<i>Mild, self-effacing, tolerant</i>	22
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>headstrong</i>	<i>headstrong</i>	<i>Headstrong</i>	<i>Gentle tempered</i>	7	<i>Wilful, egotistic, predatory</i>	<i>Mild, self-effacing, tolerant</i>	41
A/E/L	<i>Schizothymia/Dominance (Hypomania)/Paranoia</i>	<i>Cyclothymia/Submissiveness/Surgent cyclothymia</i>	<i>friendly</i>	<i>friendly</i>	<i>Hostile</i>	<i>Friendly</i>	8	<i>Rigid, tyrannical vindictive</i>	<i>Adaptable, friendly</i>	3
A/E/L	<i>Schizothymia/Dominance (Hypomania)/Paranoia</i>	<i>Cyclothymia/Submissiveness/Surgent cyclothymia</i>	<i>inflexible (emotionally)</i>	<i>inflexible</i>	<i>Inflexible (emotionally)</i>	<i>Adaptable (to change)</i>	8	<i>Rigid, tyrannical vindictive</i>	<i>Adaptable, friendly</i>	14
A/E/L	<i>Schizothymia/Dominance (Hypomania)/Paranoia</i>	<i>Cyclothymia/Submissiveness/Surgent cyclothymia</i>	<i>extra-punitive</i>	<i>severe</i>	<i>Extra-punitive</i>	-	8	<i>Rigid, tyrannical vindictive</i>	<i>Adaptable, friendly</i>	14
A/E/L	<i>Schizothymia/Dominance (Hypomania)/Paranoia</i>	<i>Cyclothymia/Submissiveness/Surgent cyclothymia</i>	<i>vindictive</i>	<i>vindictive</i>	<i>Extra-punitive</i>	-	8	<i>Rigid, tyrannical vindictive</i>	<i>Adaptable, friendly</i>	68

A /E/L	Schizothymia/ Dominance (Hypomania)/ Paranoia	Cyclothymia /Submissiveness/ Surgent cyclothymia	grateful	grateful	Thankless	Grateful	9	Surly, hard	Good- natured, easy-going	3
A /E/L	Schizothymia/ Dominance (Hypomania)/ Paranoia	Cyclothymia /Submissiveness/ Surgent cyclothymia	hard	hard	Headheart ed,embitte red	Softhearted	9	Surly, hard	Good- natured, easy-going	14
A /E/L	Schizothymia/ Dominance (Hypomania)/ Paranoia	Cyclothymia /Submissiveness/ Surgent cyclothymia	easy going	easy going	Short- tempered	Easy-going	9	Surly, hard	Good- natured, easy-going	34
C/G/J	Demoralized/D ependent, silly, incoherent/Neu rasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	idealistic	idealistic	Unrealisti c	Realistic, practical	10	Demoraliz ed, autistic	Realistic, facing life	6
C/G/J	Demoralized/D ependent, silly, incoherent/Neu rasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	abstract	abstract	Unrealisti c	Realistic, practical	10	Demoraliz ed, autistic	Realistic, facing life	7
C/G/J	Demoralized/D ependent, silly, incoherent/Neu rasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	hypothetic al	hypothetical	Unrealisti c	Realistic, practical	10	Demoraliz ed, autistic	Realistic, facing life	7
C/G/J	Demoralized/D ependent, silly, incoherent/Neu rasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	conceptual	conceptual	Unrealisti c	Realistic, practical	10	Demoraliz ed, autistic	Realistic, facing life	7
C/G/J	Demoralized/D ependent, silly, incoherent/Neu rasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	imaginary	imaginary	Unrealisti c	Realistic, practical	10	Demoraliz ed, autistic	Realistic, facing life	7
C/G/J	Demoralized/D ependent, silly, incoherent/Neu rasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	ideal	ideal	Unrealisti c	Realistic, practical	10	Demoraliz ed, autistic	Realistic, facing life	7

		<i>character</i>								
C/G/J	<i>Demoralized/Dependent, silly, incoherent/Neurasthenia</i>	<i>Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character</i>	<i>subjective</i>	<i>subjective</i>	<i>Subjective, evasive</i>	<i>Facing life</i>	10	<i>Demoralized, autistic</i>	<i>Realistic, facing life</i>	7
C/G/J	<i>Demoralized/Dependent, silly, incoherent/Neurasthenia</i>	<i>Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character</i>	<i>speculative</i>	<i>speculative</i>	<i>Unrealistic</i>	<i>Realistic, practical</i>	10	<i>Demoralized, autistic</i>	<i>Realistic, facing life</i>	13
C/G/J	<i>Demoralized/Dependent, silly, incoherent/Neurasthenia</i>	<i>Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character</i>	<i>academic</i>	<i>academic</i>	<i>Unrealistic</i>	<i>Realistic, practical</i>	10	<i>Demoralized, autistic</i>	<i>Realistic, facing life</i>	13
C/G/J	<i>Demoralized/Dependent, silly, incoherent/Neurasthenia</i>	<i>Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character</i>	<i>conjectural</i>	<i>conjectural</i>	<i>Unrealistic</i>	<i>Realistic, practical</i>	10	<i>Demoralized, autistic</i>	<i>Realistic, facing life</i>	13
C/G/J	<i>Demoralized/Dependent, silly, incoherent/Neurasthenia</i>	<i>Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character</i>	<i>persevering</i>	<i>persevering</i>	<i>Quitting</i>	<i>Persevering</i>	10	<i>Demoralized, autistic</i>	<i>Realistic, facing life</i>	32
C/G/J	<i>Demoralized/Dependent, silly, incoherent/Neurasthenia</i>	<i>Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character</i>	<i>evasive</i>	<i>evasive</i>	<i>Subjective, evasive</i>	<i>Facing life</i>	10	<i>Demoralized, autistic</i>	<i>Realistic, facing life</i>	37

B/G/J/K	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Mental Defect/Dependent, silly, incoherent/Neurasthenia/Boorishness</i>	<i>conscientious</i>	<i>conscientious</i>	<i>Conscientious</i>	<i>Conscienceless</i>	11	<i>Strong-willed, conscientious</i>	<i>Indolent, incoherent, impulsive</i>	30
B/G/J/K	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Mental Defect/Dependent, silly, incoherent/Neurasthenia/Boorishness</i>	<i>painstaking</i>	<i>painstaking</i>	<i>Painstaking</i>	<i>Slipshod</i>	11	<i>Strong-willed, conscientious</i>	<i>Indolent, incoherent, impulsive</i>	30
B/G/J/K	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Mental Defect/Dependent, silly, incoherent/Neurasthenia/Boorishness</i>	<i>persevering</i>	<i>persevering</i>	<i>Persevering</i>	<i>Quitting</i>	11	<i>Strong-willed, conscientious</i>	<i>Indolent, incoherent, impulsive</i>	32
B/K	<i>Intelligence, General Mental Capacity/Trained, socialized, cultured mind</i>	<i>Mental Defect/Boorishness</i>	<i>analytical</i>	<i>analytical</i>	<i>Analytical</i>	-	12	<i>Intellectual</i>	<i>Simple, undisciplined mind</i>	1
B/K	<i>Intelligence, General Mental Capacity/Trained, socialized, cultured mind</i>	<i>Mental Defect/Boorishness</i>	<i>thoughtful</i>	<i>thoughtful</i>	<i>Thoughtful</i>	<i>Unreflective</i>	12	<i>Intellectual</i>	<i>Simple, undisciplined mind</i>	13
B/K	<i>Intelligence, General Mental Capacity/Trained, socialized, cultured mind</i>	<i>Mental Defect/Boorishness</i>	<i>Interests wide</i>	<i>interested</i>	<i>Wide interests</i>	-	12	<i>Intellectual</i>	<i>Simple, undisciplined mind</i>	22
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>grateful</i>	<i>grateful</i>	<i>Thankless, unappreciative</i>	<i>Grateful</i>	13	<i>Insecure, infantile, hostile</i>	<i>Mature, kind, tactful</i>	3
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>jealous</i>	<i>jealous</i>	<i>Easily jealous</i>	-	13	<i>Insecure, infantile, hostile</i>	<i>Mature, kind, tactful</i>	47
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>self-pitying</i>	<i>pitying</i>	<i>Self-pitying</i>	-	13	<i>Insecure, infantile, hostile</i>	<i>Mature, kind, tactful</i>	59

A /H/K/L	Schizothymia/ Obstructive, withdrawn, schizothymia/B oorishness/Par anoia	Cyclothymia /Charitable, adventurous, cyclothymia/ Trained, socialized, cultured mind/Surgen t cyclothymia	cooperativ e	cooperative	Obstructiv e	Cooperative	14	Anti- social, schizoid	Out-going, idealistic, cooperativ e	3
A /H/K/L	Schizothymia/ Obstructive, withdrawn, schizothymia/B oorishness/Par anoia	Cyclothymia /Charitable, adventurous, cyclothymia/ Trained, socialized, cultured mind/Surgen t cyclothymia	timid (dispositio n)	timid (disposition)	Timid, withdrawn	Adventurous	14	Anti- social, schizoid	Out-going, idealistic, cooperativ e	15
A /H/K/L	Schizothymia/ Obstructive, withdrawn, schizothymia/B oorishness/Par anoia	Cyclothymia /Charitable, adventurous, cyclothymia/ Trained, socialized, cultured mind/Surgen t cyclothymia	cynical	cynical	Cynical	Idealistic	14	Anti- social, schizoid	Out-going, idealistic, cooperativ e	29
A/F/L	Schizothymia/S ugercy/Surgent cyclothymia	Cyclothymia /Agitated, melancholic desurgency/ Paranoia	genial	genial	Genial	Cold-hearted	15	Cheerful, enthusiast ic, witty	Unhappy, frustated, dour	3
A/F/L	Schizothymia/S ugercy/Surgent cyclothymia	Cyclothymia /Agitated, melancholic desurgency/ Paranoia	enthusiasti c	enthusiastic	Enthusiast ic	Apathetic	15	Cheerful, enthusiast ic, witty	Unhappy, frustated, dour	5
A/F/L	Schizothymia/S ugercy/Surgent cyclothymia	Cyclothymia /Agitated, melancholic desurgency/ Paranoia	optimistic	optimistic	Optimistic	Pessimistic	15	Cheerful, enthusiast ic, witty	Unhappy, frustated, dour	5
			impulsive temperam entally	impulsive	Impulsive	Deliberate	16	Active, neurotic, creatively unstable	Self- controlled, rigid, convention al	5
			reserved	reserved	Intrusive	Reserved	16	Active, neurotic, creatively unstable	Self- controlled, rigid, convention al	8
			hypochond rical (Catell), neurotic, hypochond riac	neurotic	Neurotic	-	16	Active, neurotic, creatively unstable	Self- controlled, rigid, convention al	53
			hypochond rical (Catell), neurotic, hypochond riac	splenetic	Neurotic	-	16	Active, neurotic, creatively unstable	Self- controlled, rigid, convention al	64

C/G	Demoralized/Dependent, silly, incoherent	Realistic, facing life/Wise, mature, polished	irritable	irritable	(Neurotic, irritable, uncontrolled)	Balanced	17	Character neurosis, psychopathic	Emotionally mature	8
C/G	Demoralized/Dependent, silly, incoherent	Realistic, facing life/Wise, mature, polished	honest	honest	Dishonest	Honest	17	Character neurosis, psychopathic	Emotionally mature	9
C/G	Demoralized/Dependent, silly, incoherent	Realistic, facing life/Wise, mature, polished	loyal (Não incluída (Não incluída na clusterização))	-	Fickle	Loyal	17	Character neurosis, psychopathic	Emotionally mature	-
			vivacious	vivacious	Vivacious	Lethargic	18	High-strung, expressive, driven	Phlegmatic	5
			high-strung	high-strung	Highly strung	(Unexcited and unexcitable)	18	High-strung, expressive, driven	Phlegmatic	8
			hurried	hurried	Hurried	Lethargic	18	High-strung, expressive, driven	Phlegmatic	42
A/H	Schizothymia/Obstructive, withdrawn, schizothymia	Cyclothymia/Charitable, adventurous, cyclothymia	cooperative	cooperative	Obstructive	Cooperative	19	Spiteful, tight-fisted, superstitious	Natural, friendly, open	3
A/H	Schizothymia/Obstructive, withdrawn, schizothymia	Cyclothymia/Charitable, adventurous, cyclothymia	friendly	friendly	Hostile	Friendly	19	Spiteful, tight-fisted, superstitious	Natural, friendly, open	3
A/H	Schizothymia/Obstructive, withdrawn, schizothymia	Cyclothymia/Charitable, adventurous, cyclothymia	frank	frank	Secretive	Frank	19	Spiteful, tight-fisted, superstitious	Natural, friendly, open	9
I	Sensitive, imaginative, anxious emotionality	Rigid, tough, poise	cheerful	cheerful	Dissatisfied	Content	20	General emotionality (with maladjustment)	Unemotional	3
I	Sensitive, imaginative, anxious emotionality	Rigid, tough, poise	emotional I(all varieties)	emotional	Emotional (in all ways)	Unemotional	20	General emotionality (with maladjustment)	Unemotional	6
I	Sensitive, imaginative, anxious emotionality	Rigid, tough, poise	excitable	excitable	Excitable	Phlegmatic	20	General emotionality (with maladjustment)	Unemotional	8

H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia /Neurasthenia	debonnaire	cosmopolitan	Debonnaire	-	21	Ascendant, expressive, widely interested	Retiring, quiet, narrow	3
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia /Neurasthenia	energetic-spirited	energetic	Energetic, spirited	Languid	21	Ascendant, expressive, widely interested	Retiring, quiet, narrow	11
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia /Neurasthenia	self-confident	self-confident	Self confident	Self distrustful	21	Ascendant, expressive, widely interested	Retiring, quiet, narrow	12
F/H/L	Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia	Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia /Paranoia	Interests special: social	amiable	Social Interests	Brooding	22	Responsive, genial, sentimental	Aloof, cold, misanthropic	3
F/H/L	Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia	Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia /Paranoia	Interests special: social	cordial	Social Interests	Brooding	22	Responsive, genial, sentimental	Aloof, cold, misanthropic	3
F/H/L	Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia	Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia /Paranoia	Interests special: social	hospitable	Social Interests	Brooding	22	Responsive, genial, sentimental	Aloof, cold, misanthropic	3
F/H/L	Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia	Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia /Paranoia	Interests special: social	informative	Social Interests	Brooding	22	Responsive, genial, sentimental	Aloof, cold, misanthropic	3
F/H/L	Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia	Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia /Paranoia	Interests special: social	social	Social Interests	Brooding	22	Responsive, genial, sentimental	Aloof, cold, misanthropic	3

F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/ Obstructive, withdrawn, schizothymia /Paranoia</i>	<i>responsive</i>	<i>responsive</i>	<i>Responsive</i>	<i>Aloof</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	3
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/ Obstructive, withdrawn, schizothymia /Paranoia</i>	<i>genial</i>	<i>genial</i>	<i>Genial</i>	<i>Coldhearted</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	3
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/ Obstructive, withdrawn, schizothymia /Paranoia</i>	<i>brooding</i>	<i>complaining</i>	<i>Social Interests</i>	<i>Brooding</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	29
			<i>eloquent</i>	<i>eloquent</i>	<i>Eloquent</i>	<i>Inarticulate</i>	23	<i>Facile, foppish, affected</i>	<i>Inarticulate, natural</i>	14
			<i>flattering</i>	<i>flattering</i>	<i>Flattering</i>	<i>Natural (not an opposite)</i>	23	<i>Facile, foppish, affected</i>	<i>Inarticulate, natural</i>	19
			<i>exhibitionistic</i>	<i>perceivable</i>	<i>Exhibitionist</i>	<i>Self-effacing</i>	23	<i>Facile, foppish, affected</i>	<i>Inarticulate, natural</i>	22
A /F/H/L	<i>Schizothymia/ Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Cyclothymia /Sugercy/Charitable, adventurous, cyclothymia/ Surgent cyclothymia</i>	<i>cautious</i>	<i>cautious</i>	<i>Suspicious</i>	<i>Trustful</i>	24	<i>Hostile, paranoid</i>	<i>Trustful, good-tempered</i>	13
A /F/H/L	<i>Schizothymia/ Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Cyclothymia /Sugercy/Charitable, adventurous, cyclothymia/ Surgent cyclothymia</i>	<i>mulish</i>	<i>mulish</i>	<i>Mulish</i>	<i>Reasonable</i>	24	<i>Hostile, paranoid</i>	<i>Trustful, good-tempered</i>	41
A /F/H/L	<i>Schizothymia/ Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Cyclothymia /Sugercy/Charitable, adventurous, cyclothymia/ Surgent cyclothymia</i>	<i>sadistic</i>	<i>inhuman, inhumane</i>	<i>Sadistic</i>	<i>Not sadistic</i>	24	<i>Hostile, paranoid</i>	<i>Trustful, good-tempered</i>	45
K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>Interests special: aesthetic (general)</i>	<i>aesthetic</i>	<i>General aesthetic interests</i>	-	25	<i>Esthetic interests, independent mind</i>	-	3

K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>independent</i>	<i>independent</i>			-	25	<i>Esthetic interests, independent mind</i>	-	43
K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>independent</i>	<i>independent</i>	<i>Independent</i>		-	25	<i>Esthetic interests, independent mind</i>	-	43
K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>independent</i>	<i>independent</i>			-	25	<i>Esthetic interests, independent mind</i>	-	43
C/D	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance</i>	<i>emotional II (sociability)</i>	<i>affable</i>	<i>Emotional</i>	<i>Unemotional</i>		26	<i>Restlessly, sthenically, hypomanically emotional</i>	<i>Calm, self-effacing, patient</i>	3
C/D	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance</i>	<i>emotional II (sociability)</i>	<i>convivial</i>	<i>Emotional</i>	<i>Unemotional</i>		26	<i>Restlessly, sthenically, hypomanically emotional</i>	<i>Calm, self-effacing, patient</i>	3
C/D	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance</i>	<i>excitable</i>	<i>excitable</i>	<i>Excitable</i>	<i>Phlegmatic</i>		26	<i>Restlessly, sthenically, hypomanically emotional</i>	<i>Calm, self-effacing, patient</i>	8
C/D	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance</i>	<i>patient</i>	<i>patient</i>	<i>Impatient (Sthenically emotional)</i>	<i>Patient</i>		26	<i>Restlessly, sthenically, hypomanically emotional</i>	<i>Calm, self-effacing, patient</i>	32
C/D/G/I	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished/Rigid, tough, poise</i>	<i>exhibitionistic</i>	<i>perceivable</i>	<i>Exhibitionist</i>	<i>Self-effacing</i>		27	<i>Infantile, demanding, self-centered</i>	<i>Emotionally mature, adjusting to frustration</i>	22
C/D/G/I	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished/Rigid, tough, poise</i>	<i>self-pitying</i>	<i>pitying</i>	<i>Self-pitying</i>		-	27	<i>Infantile, demanding, self-centered</i>	<i>Emotionally mature, adjusting to frustration</i>	59

C/D/G/I	Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality	Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished/Rigid, tough, poise	Mature (emotional)- (Não incluída na clusterização)	-	Infantile	Mature emotionally	27	Infantile, demanding, self-centered	Emotionally mature, adjusting to frustration	-
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self-respecting	self-satisfied	-	Self-respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	12
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self-respecting	self-conscious	-	Self-respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	13
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self-respecting	dignified	-	Self-respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	14
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self-respecting	elevated	-	Self-respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	14
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self-controlled	controlled	Unself-controlled	Self-controlled	28	Changeable, characterless, unrealistic	Stable, integrated character	32

B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	stable emotionality	stable	Changeable	Stable emotionally	28	Changeable, characterless, unrealistic	Stable, integrated character	61
J	Vigorous, "obsessional determined" character	Neurasthenia	alert	alert	Alert	Absent-minded	29	Psychophysically vigorous, alert	Neurasthenic	1
J	Vigorous, "obsessional determined" character	Neurasthenia	energetic-spirited	energetic	Energetic, spirited	Languid	29	Psychophysically vigorous, alert	Neurasthenic	11
J	Vigorous, "obsessional determined" character	Neurasthenia	slow	slow	Quick	Slow	29	Psychophysically vigorous, alert	Neurasthenic	63
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia/Neurasthenia	inhibited	inhibited	Incontinent	Inhibited	30	Adventurous, lusty	Generally inhibited, timid	8
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia/Neurasthenia	curious	curious	Curious	Unenquiring	30	Adventurous, lusty	Generally inhibited, timid	22
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia/Neurasthenia	gluttonous	gluttonous	Gluttonous	Queasy	30	Adventurous, lusty	Generally inhibited, timid	38
F	Sugercy	Agitated, melancholic desurgency	responsive	responsive	Responsive	Aloof	31	Sociable, hearty	Reclusive, shy	3
F	Sugercy	Agitated, melancholic desurgency	sociable II (gregarious)	gregarious	Sociable (forward, gregarious)	Shy (and reclusive)	31	Sociable, hearty	Reclusive, shy	3
F	Sugercy	Agitated, melancholic desurgency	ssociable I (meeting people)	sociable	Sociable (forward, gregarious)	Shy (and reclusive)	31	Sociable, hearty	Reclusive, shy	3
F	Sugercy	Agitated, melancholic desurgency	hearty	hearty	Hearty	Quiet	31	Sociable, hearty	Reclusive, shy	5
F	Sugercy	Agitated, melancholic desurgency	agoraphobic	reclusive	Sociable (forward, gregarious)	Shy (and reclusive)	31	Sociable, hearty	Reclusive, shy	8

F	Sugercy	Agitated, melancholic desurgency	agoraphobic	reclusive	Sociable (forward, gregarious)	Shy (and reclusive)	31	Sociable, hearty	Reclusive, shy	8
F	Sugercy	Agitated, melancholic desurgency	optimistic	optimistic	Optimistic	Pessimistic	32	-	Melancholic (agitated/revolutionary)	5
F	Sugercy	Agitated, melancholic desurgency	hypochondriacal (Catell), hypochondriac	anxious	-	Hypochondriacal	32	-	Melancholic (agitated/revolutionary)	15
F	Sugercy	Agitated, melancholic desurgency	worrying	worried	Placid	Worrying	32	-	Melancholic (agitated/revolutionary)	15
E/F/K	Dominance (Hypomania)/Sugercy/Boorishness	Submissiveness/Agitated, melancholic desurgency/ Trained, socialized, cultured mind	sensitive	sensitive	Tough	Sensitive	33	Tough, solid, talkative	Introspective, sensitive, scared	8
E/F/K	Dominance (Hypomania)/Sugercy/Boorishness	Submissiveness/Agitated, melancholic desurgency/ Trained, socialized, cultured mind	hurried	hurried	Lethargic	Hurried	33	Tough, solid, talkative	Introspective, sensitive, scared	42
E/F/K	Dominance (Hypomania)/Sugercy/Boorishness	Submissiveness/Agitated, melancholic desurgency/ Trained, socialized, cultured mind	introspective	introspective	-	Introspective	33	Tough, solid, talkative	Introspective, sensitive, scared	46
F/I	Sugercy/Sensitive, imaginative, anxious emotionality	Agitated, melancholic desurgency/ Rigid, tough, poise	intuitive	intuitive	Intuitive	Logical (precise)	34	Imaginative, introspective, constructive	Set, smug, thrifty	4
F/I	Sugercy/Sensitive, imaginative, anxious emotionality	Agitated, melancholic desurgency/ Rigid, tough, poise	Interests Special: economic	thrifty	Careless of material things	Thrifty	34	Imaginative, introspective, constructive	Set, smug, thrifty	16
F/I	Sugercy/Sensitive, imaginative, anxious emotionality	Agitated, melancholic desurgency/ Rigid, tough, poise	Interests Special: economic	frugal	Careless of material things	Thrifty	34	Imaginative, introspective, constructive	Set, smug, thrifty	16

F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>Interests Special: economic</i>	<i>sparing</i>	<i>Careless of material things</i>	<i>Thrifty</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	16
F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>Interests Special: economic</i>	<i>economical</i>	<i>Careless of material things</i>	<i>Thrifty</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	16
F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>habit-bound</i>	<i>habituated</i>	<i>Labile</i>	<i>Habit-bound</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	40
E/J/K	<i>Dominance (Hypomania)/ Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Submissiveness/Neurasthenia/Boorishness</i>	<i>Abilities: intelligence</i>	<i>intelligent</i>	<i>Intelligent</i>	<i>Stupid</i>	35	<i>Smart, assertive</i>	<i>Simple-hearted, meek</i>	1
E/J/K	<i>Dominance (Hypomania)/ Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Submissiveness/Neurasthenia/Boorishness</i>	<i>sophisticated</i>	<i>sophisticated</i>	<i>Sophisticated</i>	<i>Simple</i>	35	<i>Smart, assertive</i>	<i>Simple-hearted, meek</i>	10
E/J/K	<i>Dominance (Hypomania)/ Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Submissiveness/Neurasthenia/Boorishness</i>	<i>assertive</i>	<i>assertive</i>	<i>Assertive</i>	<i>Submissive</i>	35	<i>Smart, assertive</i>	<i>Simple-hearted, meek</i>	26

Na tabela acima, a relação entre cada traço de personalidade usado no modelo 12PF, a(s) variável(is) que pertence e seu(s) respectivo(s) fator(es) foi encontrada em CATTELL, 1946a, pp. 311-337).

Observando-se a tabela acima de imediato se nota, neste modelo, que diversos fatores abrigam variáveis comuns (e por tabela seus traços constituintes).

Como pode-se verificar, os dados da tabela acima apontam uma certa correspondência entre os *clusters* de traços definidos pelo algoritmo de clusterização MCL, seus fatores e suas variáveis constituintes.

A correspondência entre os *clusters* determinados pelo MCL e as 35 variáveis do modelo se dá no sentido de que, em várias dessas variáveis se encontram traços pertencentes

a *clusters* identificados pelo MCL, localizados em regiões próximas do grafo ou no mesmo *cluster*.

Exemplos dessas relações entre *clusters* do MCL e variáveis do modelo de 12 fatores foram detectados nas variáveis 4, 6, 8, 9, 11, 12 e 15.

A variável 4, apresenta, dentre os seus traços componentes, dois traços (*thoughtful* e *austere*) pertencentes a *clusters* diferentes (13 e 14, respectivamente), porém localizados em regiões próximas do grafo.

A variável 6 é composta por três traços de personalidade; dois deles (*grateful* e *friendly*) sendo identificados no mesmo *cluster* (*cluster* 3) segundo o algoritmo MCL .

A variável 8 contém, dentre os seus traços componentes, dois traços (*inflexible* e *severe*) pertencentes a um mesmo *cluster* (14, identificado pelo MCL) e mais um outro traço (*friendly*), pertencente ao *cluster* 3, localizado em região do grafo próxima a do *cluster* 14.

A variável 9, contém os traços *grateful* e *hard*, pertencentes respectivamente aos *clusters* 3 e 14 (segundo o MCL), próximos entre si no grafo.

A variável 11, contém os traços *conscientious* e *painstaking*, ambos localizados no mesmo *cluster* 30, identificado pelo MCL.

A variável 12 contém os traços *analytical* e *thoughtful*, pertencentes respectivamente aos *clusters* 1 e 13 (identificados segundo o MCL). Estes *clusters* são vizinhos no grafo representando a rede de relações entre os traços selecionados.

Finalmente a variável 15 é composta por traços pertencentes aos *clusters* 3 e 5, identificados pelo MCL. Estes *clusters* são vizinhos no grafo associado às ligações entre os 222 termos selecionados.

No que se refere aos padrões de relacionamento observados entre traços de personalidade pertencentes aos diferentes *clusters* (identificados pelo MCL) e os fatores do modelo, observou-se que os fatores que indiretamente abrigam traços de personalidade que foram identificados nos mesmos *clusters* (pelo algoritmo MCL) ou em *clusters* uns dos outros no grafo, são os que abrigam as variáveis nas quais se observou estes padrões.

Assim, por exemplo, nos fatores H, I e L, que abrigam a variável 6, observou-se dois traços (*grateful* e *friendly*) alocados em um mesmo *cluster* pelo algoritmo MCL, pois este foi o padrão observado nesta variável.

Similarmente, nos fatores A, E e L observou-se a presença dos traços *friendly*, *inflexible* e *severe*. Os dois últimos foram identificados no mesmo *cluster* (14) pelo algoritmo MCL. O primeiro, foi alocado pelo MCL ao *cluster* 3, localizado em região do

grafo próxima à região onde se localiza o *cluster* 14. Pôde-se tirar estas conclusões porque observou-se que estes fatores (A, E e L) abrigam a variável 8, que por sua vez é composta pelos traços acima mencionados; e assim sucessivamente para todos os demais fatores contendo variáveis onde tais padrões estão presentes.

Assim, nesses sentidos explicitados acima pôde-se verificar que os *clusters* de palavras, organizadas de acordo com seus relacionamentos, segundo a linguagem natural, não conseguiram captar de modo satisfatório a composição das variáveis e dos fatores do modelo de Cattell.

De modo geral, 40% (14) das variáveis do modelo tiveram pelo menos 2 de seus traços (ou dos traços e elas correspondentes, no caso de traços candidatos a substitutos de traços originalmente propostos por Cattell e ausentes no TTT) alocados em um mesmo *cluster* pelo MCL.

Assim, 60% das variáveis do modelo de Cattell corresponderam a traços alocados a clusters diferentes pelo MCL.

Tabela 34: Comparação dos fatores do modelo 12 PF de Cattell (1945b), com os clusters obtidos com o algoritmo SOM.

Fatores do modelo 12PF (CATTELL 1945) contendo as 35 variáveis	Descrição dos polos do fator		Traço correspondente na seleção inicial de 171 (CATTELL, 1943b)	Traço correspondente utilizado neste trabalho	Polos dos traços pertencentes ao modelo 12PF		Número da Variável (Cluster) contendo o traço (Cattell - 1945)	Polos da Variável contendo o traço		Clusters SOM
A/H	<i>Schizothymia/Obstructive, withdrawn, schizothymia</i>	<i>Cyclothymia/Charitable, adventurous, cyclothymia</i>	<i>frank</i>	<i>frank</i>	<i>Secretive</i>	<i>Frank</i>	19	<i>Spiteful, tight-fisted, superstitious</i>	<i>Natural, friendly, open</i>	1
C/G	<i>Demoralized/Dependent, silly, incoherent</i>	<i>Realistic, facing life/Wise, mature, polished</i>	<i>honest</i>	<i>honest</i>	<i>Dishonest</i>	<i>Honest</i>	17	<i>Character neurosis, psychopathic</i>	<i>Emotionally mature</i>	2
B/J	<i>Intelligence, General Mental Capacity/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Neurasthenia</i>	<i>clever</i>	<i>clever</i>	<i>Clever</i>	-	2	<i>Intelligent, analytical</i>	<i>Unimaginative, stupid</i>	3
J	<i>Vigorous, "obsessional determined" character</i>	<i>Neurasthenia</i>	<i>alert</i>	<i>alert</i>	<i>Alert</i>	<i>Absent-minded</i>	29	<i>Psychophysically vigorous, alert</i>	<i>Neurasthenic</i>	5

F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>Interests Special: economic</i>	<i>thrifty</i>	<i>Careless of material things</i>	<i>Thrifty</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	6
F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>Interests Special: economic</i>	<i>frugal</i>	<i>Careless of material things</i>	<i>Thrifty</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	6
F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>Interests Special: economic</i>	<i>sparing</i>	<i>Careless of material things</i>	<i>Thrifty</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	6
F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>Interests Special: economic</i>	<i>economical</i>	<i>Careless of material things</i>	<i>Thrifty</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	6
K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>Interests special: aesthetic (general)</i>	<i>aesthetic</i>	<i>General aesthetic interests</i>	-	25	<i>Esthetic interests, independent mind</i>	-	7
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Interests special: social</i>	<i>hospitable</i>	<i>Social Interests</i>	<i>Brooding</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	7
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Interests special: social</i>	<i>informative</i>	<i>Social Interests</i>	<i>Brooding</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	7
I	<i>Sensitive, imaginative, anxious emotionality</i>	<i>Rigid, tough, poise</i>	<i>cheerful</i>	<i>cheerful</i>	<i>Dissatisfied</i>	<i>Content</i>	20	<i>General emotionality (with maladjustment)</i>	<i>Unemotional</i>	8
B/J	<i>Intelligence, General Mental Capacity/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Neurasthenia</i>	<i>clear thinking</i>	<i>coherent</i>	<i>Clear Thinking</i>	<i>Incoherent, confused</i>	2	<i>Intelligent, analytical</i>	<i>Unimaginative, stupid</i>	8

A /H/K/L	Schizothymia/ Obstructive, withdrawn, schizothymia/ Boorishness/P aranoia	Cyclothymia /Charitable, adventurous, cyclothymia/ Trained, socialized, cultured mind/Surgen t cyclothymia	cooperative	cooperative	Obstructiv e	Cooperati ve	14	Anti-social, schizoid	Out-going, idealistic, cooperative	8
A /H	Schizothymia/ Obstructive, withdrawn, schizothymia	Cyclothymia /Charitable, adventurous, cyclothymia	cooperative	cooperative	Obstructiv e	Cooperati ve	19	Spiteful, tight-fisted, superstitious	Natural, friendly, open	8
H/J	Charitable, adventurous, cyclothymia/V igorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia /Neurastheni a	debonnaire	cosmopolita n	Debonnair e	-	21	Ascendant, expressive, widely interested	Retiring, quiet, narrow	8
C/D	Demoralized/ Hypersensitiv e, infantile, sthenic emotionality	Realistic, facing life/Phlegma tic frustration tolerance	emotional II (sociability)	affable	Emotional	Unemotio nal	26	Restlessly, sthenically, hypomanical ly emotional	Calm, self- effacing, patient	8
C/D	Demoralized/ Hypersensitiv e, infantile, sthenic emotionality	Realistic, facing life/Phlegma tic frustration tolerance	emotional II (sociability)	convivial	Emotional	Unemotio nal	26	Restlessly, sthenically, hypomanical ly emotional	Calm, self- effacing, patient	8
A /E/L	Schizothymia/ Dominance (Hypomania)/ Paranoia	Cyclothymia /Submissiven ess/Surgent cyclothymia	friendly	friendly	Hostile	Friendly	8	Rigid, tyrannical, vindictive	Adaptable, friendly	8
A /H	Schizothymia/ Obstructive, withdrawn, schizothymia	Cyclothymia /Charitable, adventurous, cyclothymia	friendly	friendly	Hostile	Friendly	19	Spiteful, tight-fisted, superstitious	Natural, friendly, open	8
A /F/L	Schizothymia/ Sugercy/Surge nt cyclothymia	Cyclothymia /Agitated, melancholic desurgency/ Paranoia	genial	genial	Genial	Cold- hearted	15	Cheerful, enthusiastic, witty	Unhappy, frustated, dour	8
H/VL	Obstructive, withdrawn, schizothymia/ Rigid, tough, poise/Paranoi a	Charitable, adventurous, cyclothymia/ Sensitive, imaginative, anxious emotionality/ Surgent cyclothymia	grateful	grateful	Thankless	Grateful	6	Hard, cynical	Kindly, gentle, idealistic	8

F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Interests special: social</i>	<i>amiable</i>	<i>Social Interests</i>	<i>Brooding</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	8
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Interests special: social</i>	<i>cordial</i>	<i>Social Interests</i>	<i>Brooding</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	8
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>Interests special: social</i>	<i>social</i>	<i>Social Interests</i>	<i>Brooding</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	8
H/L	<i>Obstructive, withdrawn, schizothymia/Rigid, tough, poise/Paranoia</i>	<i>Charitable, adventurous, cyclothymia/Sensitive, imaginative, anxious emotionality/Surgent cyclothymia</i>	<i>friendly</i>	<i>friendly</i>	<i>Hostile</i>	<i>Friendly, understanding</i>	6	<i>Hard, cynical</i>	<i>Kindly, gentle, idealistic</i>	8
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>responsive</i>	<i>responsive</i>	<i>Responsive</i>	<i>Aloof</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	8
F	<i>Sugercy</i>	<i>Agitated, melancholic desurgency</i>	<i>responsive</i>	<i>responsive</i>	<i>Responsive</i>	<i>Aloof</i>	31	<i>Sociable, hearty</i>	<i>Reclusive, shy</i>	8
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia</i>	<i>genial</i>	<i>genial</i>	<i>Genial</i>	<i>Coldhearted</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	8
F	<i>Sugercy</i>	<i>Agitated, melancholic desurgency</i>	<i>sociable II (gregarious)</i>	<i>gregarious</i>	<i>Sociable (forward, gregarious)</i>	<i>Shy (and reclusive)</i>	31	<i>Sociable, hearty</i>	<i>Reclusive, shy</i>	8
F	<i>Sugercy</i>	<i>Agitated, melancholic desurgency</i>	<i>ssociable I (meeting people)</i>	<i>sociable</i>	<i>Sociable (forward, gregarious)</i>	<i>Shy (and reclusive)</i>	31	<i>Sociable, hearty</i>	<i>Reclusive, shy</i>	8
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>grateful</i>	<i>grateful</i>	<i>Thankless, unappreciative</i>	<i>Grateful</i>	13	<i>Insecure, infantile, hostile</i>	<i>Mature, kind, tactful</i>	8

F	Sugercy	Agitated, melancholic desurgency	optimistic	optimistic	Optimistic	Pessimistic	32	-	Melancholic (agitated/involuntional)	8
A /F/L	Schizothymia/Sugercy/Surgent cyclothymia	Cyclothymia /Agitated, melancholic desurgency/Paranoia	optimistic	optimistic	Optimistic	Pessimistic	15	Cheerful, enthusiastic, witty	Unhappy, frustated, dour	8
A /E/L	Schizothymia/Dominance (Hypomania)/Paranoia	Cyclothymia /Submissiveness/Surgent cyclothymia	grateful	grateful	Thankless	Grateful	9	Surly, hard	Good-natured, easy-going	8
A /E/L	Schizothymia/Dominance (Hypomania)/Paranoia	Cyclothymia /Submissiveness/Surgent cyclothymia	vindictive	vindictive	Extra-punitive	-	8	Rigid, tyrannical, vindictive	Adaptable, friendly	11
D/E	Hypersensitive, infantile, sthenic emotionality/Dominance (Hypomania)	Phlegmatic frustration tolerance/Submissiveness	assertive	assertive	Assertive	Submissive	1	Self Assertive	Self Submissive	14
E/J/K	Dominance (Hypomania)/Vigorous, "obsessional determined" character/Trained, socialized, cultured mind	Submissiveness/Neurasthenia/Boorishness	assertive	assertive	Assertive	Submissive	35	Smart, assertive	Simple-hearted, meek	14
B/G/J/K	Intelligence, General Mental Capacity/Wise, mature, polished/Vigorous, "obsessional determined" character/Trained, socialized, cultured mind	Mental Defect/Dependent, silly, incoherent/Neurasthenia/Boorishness	conscientious	conscientious	Conscientious	Conscientless	11	Strong-willed, conscientious	Indolent, incoherent, impulsive	15
B/G/J/K	Intelligence, General Mental Capacity/Wise, mature, polished/Vigorous, "obsessional determined" character/Trained, socialized, cultured mind	Mental Defect/Dependent, silly, incoherent/Neurasthenia/Boorishness	painstaking	painstaking	Painstaking	Slipshod	11	Strong-willed, conscientious	Indolent, incoherent, impulsive	15
F	Sugercy	Agitated, melancholic desurgency	hearty	hearty	Hearty	Quiet	31	Sociable, hearty	Reclusive, shy	16
B/K	Intelligence, General Mental Capacity/Trained, socialized, cultured mind	Mental Defect/Boorishness	analytical	analytical	Analytical	-	12	Intellectual	Simple, undisciplined mind	17

B/J	<i>Intelligence, General Mental Capacity/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Neurasthenia</i>	<i>Abilities: intelligence</i>	<i>intelligent</i>	<i>Intelligent</i>	<i>Stupid</i>	2	<i>Intelligent, analytical</i>	<i>Unimaginative, stupid</i>	18
E/J/K	<i>Dominance (Hypomania)/ Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Submissiveness/Neurasthenia/Boorishness</i>	<i>Abilities: intelligence</i>	<i>intelligent</i>	<i>Intelligent</i>	<i>Stupid</i>	35	<i>Smart, assertive</i>	<i>Simple-hearted, meek</i>	18
H/V/L	<i>Obstructive, withdrawn, schizothymia/ Rigid, tough, poise/Paranoia</i>	<i>Charitable, adventurous, cyclothymia/ Sensitive, imaginative, anxious emotionality/ Surgent cyclothymia</i>	<i>hard</i>	<i>hard</i>	<i>Hardhearted</i>	<i>Softhearted</i>	6	<i>Hard, cynical</i>	<i>Kindly, gentle, idealistic</i>	19
A /E/L	<i>Schizothymia/ Dominance (Hypomania)/ Paranoia</i>	<i>Cyclothymia /Submissiveness/Surgent cyclothymia</i>	<i>hard</i>	<i>hard</i>	<i>Headhearted, embittered</i>	<i>Softhearted</i>	9	<i>Surly, hard</i>	<i>Good-natured, easy-going</i>	19
E/J/K	<i>Dominance (Hypomania)/ Vigorous, "obsessional determined" character/Trained, socialized, cultured mind</i>	<i>Submissiveness/Neurasthenia/Boorishness</i>	<i>sophisticated</i>	<i>sophisticated</i>	<i>Sophisticated</i>	<i>Simple</i>	35	<i>Smart, assertive</i>	<i>Simple-hearted, meek</i>	20
F/H/L	<i>Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>Agitated, melancholic desurgency/ Obstructive, withdrawn, schizothymia /Paranoia</i>	<i>brooding</i>	<i>complaining</i>	<i>Social Interests</i>	<i>Brooding</i>	22	<i>Responsive, genial, sentimental</i>	<i>Aloof, cold, misanthropic</i>	28
A /H/K/L	<i>Schizothymia/ Obstructive, withdrawn, schizothymia/ Boorishness/Paranoia</i>	<i>Cyclothymia /Charitable, adventurous, cyclothymia/ Trained, socialized, cultured mind/Surgent cyclothymia</i>	<i>cynical</i>	<i>cynical</i>	<i>Cynical</i>	<i>Idealistic</i>	14	<i>Anti-social, schizoid</i>	<i>Out-going, idealistic, cooperative</i>	28
H/J	<i>Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character</i>	<i>Obstructive, withdrawn, schizothymia /Neurasthenia</i>	<i>energetic-spirited</i>	<i>energetic</i>	<i>Energetic, spirited</i>	<i>Languid</i>	21	<i>Ascendant, expressive, widely interested</i>	<i>Retiring, quiet, narrow</i>	29

J	Vigorous, "obsessional determined" character	Neurasthenia	energetic-spirited	energetic	Energetic, spirited	Languid	29	Psychophysically vigorous, alert	Neurasthenic	29
A/E/L	Schizothymia/Dominance (Hypomania)/Paranoia	Cyclothymia/Submissiveness/Surgent cyclothymia	inflexible (emotionally)	inflexible	Inflexible (emotionally)	Adaptable (to change)	8	Rigid, tyrannical, vindictive	Adaptable, friendly	32
A/E/L	Schizothymia/Dominance (Hypomania)/Paranoia	Cyclothymia/Submissiveness/Surgent cyclothymia	extra-punitive	severe	Extra-punitive	-	8	Rigid, tyrannical, vindictive	Adaptable, friendly	33
E	Dominance (Hypomania)	Submissiveness	extra-punitive	severe	Extra-punitive (blaming mistakes on others)	-	7	Wilful, egotistic, predatory	Mild, self-effacing, tolerant	33
B/C/G	Mental Defect/Demoralized/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Wise, mature, polished	austere	austere	Profligate	Austere	4	Changeable, frivolous	Thoughtful, stoic, reserved	34
A/E/L	Schizothymia/Dominance (Hypomania)/Paranoia	Cyclothymia/Submissiveness/Surgent cyclothymia	easy going	easy going	Short-tempered	Easy-going	9	Surly, hard	Good-natured, easy-going	36
E/F/K	Dominance (Hypomania)/Sugercy/Boorishness	Submissiveness/Agitated, melancholic desurgency/Trained, socialized, cultured mind	introspective	introspective	-	Introspective	33	Tough, solid, talkative	Introspective, sensitive, scared	37
F/I	Sugercy/Sensitive, imaginative, anxious emotionality	Agitated, melancholic desurgency/Rigid, tough, poise	habit-bound	habituated	Labile	Habit-bound	34	Imaginative, introspective, constructive	Set, smug, thrifty	39
C/D/G/I	Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality	Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished/Rigid, tough, poise	self-pitying	pitying	Self-pitying	-	27	Infantile, demanding, self-centered	Emotionally mature, adjusting to frustration	41
E	Dominance (Hypomania)	Submissiveness	self-pitying	pitying	Self-pitying	-	13	Insecure, infantile, hostile	Mature, kind, tactful	41

B/G/I/J	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Rigid, tough, poise/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality/Neurasthenia</i>	<i>reliable</i>	<i>reliable</i>	<i>Reliable</i>	<i>Undependable</i>	3	<i>Wise, mature, polished</i>	<i>Dependent, silly, incoherent</i>	43
B/C/D/G	<i>Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent</i>	<i>Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished</i>	<i>stable emotionally</i>	<i>stable</i>	<i>Changeable</i>	<i>Stable emotionally</i>	28	<i>Changeable, characterless, unrealistic</i>	<i>Stable, integrated character</i>	43
C/D	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance</i>	<i>patient</i>	<i>patient</i>	<i>Impatient (Sihencally emotional)</i>	<i>Patient</i>	26	<i>Restlessly, sthenically, hypomanically emotional</i>	<i>Calm, self-effacing, patient</i>	44
B/C/D/G	<i>Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent</i>	<i>Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished</i>	<i>self-controlled</i>	<i>controlled</i>	<i>Unself-controlled</i>	<i>Self-controlled</i>	28	<i>Changeable, characterless, unrealistic</i>	<i>Stable, integrated character</i>	44
A/F/H/L	<i>Schizothymia/Agitated, melancholic desurgency/Obsstructive, withdrawn, schizothymia/Paranoia</i>	<i>Cyclothymia/Sugercy/Charitable, adventurous, cyclothymia/Surgent cyclothymia</i>	<i>cautious</i>	<i>cautious</i>	<i>Suspicious</i>	<i>Trustful</i>	24	<i>Hostile, paranoid</i>	<i>Trustful, good-tempered</i>	45
B/C/G	<i>Mental Defect/Demoralized/Dependent, silly, incoherent</i>	<i>Intelligence, General Mental Capacity/Realistic, facing life/Wise, mature, polished</i>	<i>thoughtful</i>	<i>thoughtful</i>	<i>Unreflective</i>	<i>Thoughtful</i>	4	<i>Changeable, frivolous</i>	<i>Thoughtful, stoic, reserved</i>	45
B/K	<i>Intelligence, General Mental Capacity/Trained, socialized, cultured mind</i>	<i>Mental Defect/Boorishness</i>	<i>thoughtful</i>	<i>thoughtful</i>	<i>Thoughtful</i>	<i>Unreflective</i>	12	<i>Intellectual</i>	<i>Simple, undisciplined mind</i>	45

C/G/J	Demoralized/ Dependent, silly, incoherent/Neurasthenia	Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character	hypothetical	hypothetical	Unrealistic	Realistic, practical	10	Demoralized, autistic	Realistic, facing life	46
C/G/J	Demoralized/ Dependent, silly, incoherent/Neurasthenia	Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character	ideal	ideal	Unrealistic	Realistic, practical	10	Demoralized, autistic	Realistic, facing life	46
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia /Neurasthenia	inhibited	inhibited	Incontinent	Inhibited	30	Adventurous, lustful	Generally inhibited, timid	47
			reserved	reserved	Intrusive	Reserved	16	Active, neurotic, creatively unstable	Self- controlled, rigid, conventional	47
F	Sugercy	Agitated, melancholic desurgency	agoraphobic	reclusive	Sociable (forward, gregarious)	Shy (and reclusive)	31	Sociable, hearty	Reclusive, shy	47
F	Sugercy	Agitated, melancholic desurgency	agoraphobic	reclusive	Sociable (forward, gregarious)	Shy (and reclusive)	31	Sociable, hearty	Reclusive, shy	47
			eloquent	eloquent	Eloquent	Inarticulate	23	Facile, foppish, affected	Inarticulate, natural	48
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/ Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self- respecting	dignified	-	Self- respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	48
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/ Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self- respecting	elevated	-	Self- respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	48

A /F/H/L	Schizothymia/ Agitated, melancholic desurgency/O bstructive, withdrawn, schizothymia/ Paranoia	Cyclothymia /Sugercy/Ch aritable, adventurous, cyclothymia/ Surgent cyclothymia	sadistic	inhuman, inhumane	Sadistic	Not sadistic	24	Hostile, paranoid	Trustful, good- tempered	50
J	Vigorous, "obsessional determined" character	Neurasthenia	slow	slow	Quick	Slow	29	Psychophysically vigorous, alert	Neurasthenic	53
			hypochondriacal (Catell), neurotic, hypochondriac	splenetic	Neurotic	-	16	Active, neurotic, creatively unstable	Self- controlled, rigid, conventional	57
C/G/J	Demoralized/ Dependent, silly, incoherent/Neurasthenia	Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character	persevering	persevering	Quitting	Persevering	10	Demoralized, autistic	Realistic, facing life	58
B/G/J/K	Intelligence, General Mental Capacity/Wise, mature, polished/Vigorous, "obsessional determined" character/Trained, socialized, cultured mind	Mental Defect/Dependent, silly, incoherent/Neurasthenia/ Boorishness	persevering	persevering	Persevering	Quitting	11	Strong-willed, conscientious	Indolent, incoherent, impulsive	58
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/ Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished	self-respecting	self-conscious	-	Self-respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	61
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia/ Neurasthenia	self-confident	self-confident	Self-confident	Self-distrustful	21	Ascendant, expressive, widely interested	Retiring, quiet, narrow	62
B/C/D/G	Mental Defect/Demoralized/Hypersensitive, infantile, sthenic emotionality/ Dependent, silly,	Intelligence, General Mental Capacity/Realistic, facing life/Phlegmatic frustration	self-respecting	self-satisfied	-	Self-respecting	28	Changeable, characterless, unrealistic	Stable, integrated character	62

	<i>incoherent</i>	<i>tolerance/Wise, mature, polished</i>								
K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>independent</i>	<i>independent</i>		-	25	<i>Esthetic interests, independent mind</i>	-	63
B/G/I/J	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Rigid, tough, poise/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality/Neurasthenia</i>	<i>independent</i>	<i>independent</i>	<i>Independent</i>	<i>Emotionally dependent</i>	3	<i>Wise, mature, polished</i>	<i>Dependent, silly, incoherent</i>	63
K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>independent</i>	<i>independent</i>	<i>Independent</i>	-	25	<i>Esthetic interests, independent mind</i>	-	63
K	<i>Trained, socialized, cultured mind</i>	<i>Boorishness</i>	<i>independent</i>	<i>independent</i>		-	25	<i>Esthetic interests, independent mind</i>	-	63
			<i>hurried</i>	<i>hurried</i>	<i>Hurried</i>	<i>Lethargic</i>	18	<i>High-strung, expressive, driven</i>	<i>Phlegmatic</i>	67
E/F/K	<i>Dominance (Hypomania)/Sugercy/Boorishness</i>	<i>Submissiveness/Agitated, melancholic desurgency/Trained, socialized, cultured mind</i>	<i>hurried</i>	<i>hurried</i>	<i>Lethargic</i>	<i>Hurried</i>	33	<i>Tough, solid, talkative</i>	<i>Introspective, sensitive, scared</i>	67
C/G/J	<i>Demoralized/Dependent, silly, incoherent/Neurasthenia</i>	<i>Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character</i>	<i>evasive</i>	<i>evasive</i>	<i>Subjective, evasive</i>	<i>Facing life</i>	10	<i>Demoralized, autistic</i>	<i>Realistic, facing life</i>	68
F/I	<i>Sugercy/Sensitive, imaginative, anxious emotionality</i>	<i>Agitated, melancholic desurgency/Rigid, tough, poise</i>	<i>intuitive</i>	<i>intuitive</i>	<i>Intuitive</i>	<i>Logical (precise)</i>	34	<i>Imaginative, introspective, constructive</i>	<i>Set, smug, thrifty</i>	71

C/G/J	Demoralized/Dependent, silly, incoherent/Neurasthenia	Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character	speculative	speculative	Unrealistic	Realistic, practical	10	Demoralized, autistic	Realistic, facing life	72
C/G/J	Demoralized/Dependent, silly, incoherent/Neurasthenia	Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character	conjectural	conjectural	Unrealistic	Realistic, practical	10	Demoralized, autistic	Realistic, facing life	72
C/G/J	Demoralized/Dependent, silly, incoherent/Neurasthenia	Realistic, facing life/Wise, mature, polished/Vigorous, "obsessional determined" character	academic	academic	Unrealistic	Realistic, practical	10	Demoralized, autistic	Realistic, facing life	73
D/E	Hypersensitive, infantile, sthenic emotionality/Dominance (Hypomania)	Phlegmatic frustration tolerance/Submissiveness	boastful	boastful	Boastful	Modest	1	Self Assertive	Self Submissive	74
D/E	Hypersensitive, infantile, sthenic emotionality/Dominance (Hypomania)	Phlegmatic frustration tolerance/Submissiveness	conceited	conceited	Conceited	Self critical, dissatisfied	1	Self Assertive	Self Submissive	75
			flattering	flattering	Flattering	Natural (not an opposite)	23	Facile, foppish, affected	Inarticulate, natural	76
B/C/G	Mental Defect/Demoralized/Dependent, silly, incoherent	Intelligence, General Mental Capacity/Realistic, facing life/Wise, mature, polished	impulsive temperament ally	impulsive	Impulsive	Deliberate	4	Changeable, frivolous	Thoughtful, stoic, reserved	82
			impulsive temperament ally	impulsive	Impulsive	Deliberate	16	Active, neurotic, creatively unstable	Self-controlled, rigid, conventional	82
I	Sensitive, imaginative, anxious emotionality	Rigid, tough, poised	emotional I(all varieties)	emotional	Emotional (in all ways)	Unemotional	20	General emotionality (with maladjustment)	Unemotional	83

C/G/J	Demoralized/ Dependent, silly, incoherent/Ne urasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	idealistic	idealistic	Unrealisti c	Realistic, practical	10	Demoralized , autistic	Realistic, facing life	84
C/G/J	Demoralized/ Dependent, silly, incoherent/Ne urasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	imaginary	imaginary	Unrealisti c	Realistic, practical	10	Demoralized , autistic	Realistic, facing life	85
C/G/J	Demoralized/ Dependent, silly, incoherent/Ne urasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	subjective	subjective	Subjective, evasive	Facing life	10	Demoralized , autistic	Realistic, facing life	85
C/G/J	Demoralized/ Dependent, silly, incoherent/Ne urasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	abstract	abstract	Unrealisti c	Realistic, practical	10	Demoralized , autistic	Realistic, facing life	86
			vivacious	vivacious	Vivacious	Lethargic	18	High-strung, expressive, driven	Phlegmatic	94
C/G/J	Demoralized/ Dependent, silly, incoherent/Ne urasthenia	Realistic, facing life/Wise, mature, polished/Vig orous, "obsessiona l determined" character	conceptual	conceptual	Unrealisti c	Realistic, practical	10	Demoralized , autistic	Realistic, facing life	96
F	Sugercy	Agitated, melancholic desurgency	worrying	worried	Placid	Worrying	32	-	Melancholi c (agitated/in voluntional)	97
E	Dominance (Hypomania)	Submissiven ess	jealous	jealous	Easly jealous	-	13	Insecure, infantile, hostile	Mature, kind, tactful	98
E	Dominance (Hypomania)	Submissiven ess	headstrong	headstrong	Headstron g	Gentle tempered	7	Wilful, egotistic, predatory	Mild, self- effacing, tolerant	99

A /F/H/L	Schizothymia/ Agitated, melancholic desurgency/Obstructive, withdrawn, schizothymia/Paranoia	Cyclothymia /Sugercy/Charitable, adventurous, cyclothymia/ Surgent cyclothymia	mulish	mulish	Mulish	Reasonable	24	Hostile, paranoid	Trustful, good-tempered	99
D/I	Hypersensitive, infantile, sthenic emotionality/Sensitive, imaginative, anxious emotionality	Phlegmatic frustration tolerance/Rigid, tough, poise	hypochondriacal (Catell), neurotic, hypochondriac	neurotic	Nervous, especific neurotic symptoms	-	5	Neurotic	Not generally neurotic	101
			hypochondriacal (Catell), neurotic, hypochondriac	neurotic	Neurotic	-	16	Active, neurotic, creatively unstable	Self-controlled, rigid, conventional	101
H/J	Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character	Obstructive, withdrawn, schizothymia /Neurasthenia	curious	curious	Curious	Unenquiring	30	Adventurous, lusty	Generally inhibited, timid	102
C/G	Demoralized/Dependent, silly, incoherent	Realistic, facing life/Wise, mature, polished	irritable	irritable	(Neurotic, irritable, uncontrolled)	Balanced	17	Character neurosis, psychopathic	Emotionally mature	104
A /F/L	Schizothymia/ Sugercy/Surgent cyclothymia	Cyclothymia /Agitated, melancholic desurgency/Paranoia	enthusiastic	enthusiastic	Enthusiastic	Apathetic	15	Cheerful, enthusiastic, witty	Unhappy, frustated, dour	105
I	Sensitive, imaginative, anxious emotionality	Rigid, tough, poise	excitable	excitable	Excitable	Phlegmatic	20	General emotionality (with maladjustment)	Unemotional	106
C/D	Demoralized/ Hypersensitive, infantile, sthenic emotionality	Realistic, facing life/Phlegmatic frustration tolerance	excitable	excitable	Excitable	Phlegmatic	26	Restlessly, sthenically, hypomanically emotional	Calm, self-effacing, patient	106
A /H/K/L	Schizothymia/ Obstructive, withdrawn, schizothymia/ Boorishness/Paranoia	Cyclothymia /Charitable, adventurous, cyclothymia/ Trained, socialized, cultured mind/Surgent cyclothymia	timid (disposition)	timid (disposition)	Timid, withdrawn	Adventurous	14	Anti-social, schizoid	Out-going, idealistic, cooperative	108

H/J	<i>Charitable, adventurous, cyclothymia/Vigorous, "obsessional determined" character</i>	<i>Obstructive, withdrawn, schizothymia /Neurasthenia</i>	<i>gluttonous</i>	<i>gluttonous</i>	<i>Gluttonous</i>	<i>Queasy</i>	30	<i>Adventurous, lusty</i>	<i>Generally inhibited, timid</i>	110
B/K	<i>Intelligence, General Mental Capacity/Trained, socialized, cultured mind</i>	<i>Mental Defect/Boorishness</i>	<i>Interests wide</i>	<i>interested</i>	<i>Wide interests</i>	-	12	<i>Intellectual</i>	<i>Simple, undisciplined mind</i>	113
E/F/K	<i>Dominance (Hypomania)/Sugercy/Boorishness</i>	<i>Submissiveness/Agitated, melancholic desurgency/Trained, socialized, cultured mind</i>	<i>sensitive</i>	<i>sensitive</i>	<i>Tough</i>	<i>Sensitive</i>	33	<i>Tough, solid, talkative</i>	<i>Introspective, sensitive, scared</i>	115
			<i>high-strung</i>	<i>high-strung</i>	<i>Highly strung</i>	<i>(Unexcited and unexcitable)</i>	18	<i>High-strung, expressive, driven</i>	<i>Phlegmatic</i>	116
D/I	<i>Hypersensitive, infantile, sthenic emotionality/Sensitive, imaginative, anxious emotionality</i>	<i>Phlegmatic frustration tolerance/Rigid, tough, poise</i>	<i>hypochondriacal (Catell), hypochondriac</i>	<i>anxious</i>	<i>Hypochondriacal</i>	-	5	<i>Neurotic</i>	<i>Not generally neurotic</i>	117
F	<i>Sugercy</i>	<i>Agitated, melancholic desurgency</i>	<i>hypochondriacal (Catell), hypochondriac</i>	<i>anxious</i>	-	<i>Hypochondriacal</i>	32	-	<i>Melancholic (agitated/involuntal)</i>	117
E	<i>Dominance (Hypomania)</i>	<i>Submissiveness</i>	<i>exhibitionistic</i>	<i>perceivable</i>	<i>Exhibitionist</i>	<i>Self-effacing</i>	7	<i>Wilful, egotistic, predatory</i>	<i>Mild, self-effacing, tolerant</i>	122
			<i>exhibitionistic</i>	<i>perceivable</i>	<i>Exhibitionist</i>	<i>Self-effacing</i>	23	<i>Facile, foppish, affected</i>	<i>Inarticulate, natural</i>	122
C/D/G/I	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished/Rigid, tough, poise</i>	<i>exhibitionistic</i>	<i>perceivable</i>	<i>Exhibitionist</i>	<i>Self-effacing</i>	27	<i>Infantile, demanding, self-centered</i>	<i>Emotionally mature, adjusting to frustration</i>	122
C/G	<i>Demoralized/Dependent, silly, incoherent</i>	<i>Realistic, facing life/Wise, mature, polished</i>	<i>loyal (Não incluída(Não incluída na clusterização))</i>	-	<i>Fickle</i>	<i>Loyal</i>	17	<i>Character neurosis, psychopathic</i>	<i>Emotionally mature</i>	-

C/D/G/I	<i>Demoralized/Hypersensitive, infantile, sthenic emotionality/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality</i>	<i>Realistic, facing life/Phlegmatic frustration tolerance/Wise, mature, polished/Rigid, tough, poise</i>	<i>Mature (emotional)-(Não incluída na clusterização)</i>	-	<i>Infantile</i>	<i>Mature emotionally</i>	27	<i>Infantile, demanding, self-centered</i>	<i>Emotionally mature, adjusting to frustration</i>	-
B/G/I/J	<i>Intelligence, General Mental Capacity/Wise, mature, polished/Rigid, tough, poise/Vigorous, "obsessional determined" character</i>	<i>Mental Defect/Dependent, silly, incoherent/Sensitive, imaginative, anxious emotionality/Neurasthenia</i>	<i>Mature (emotional)-(Não incluída)(Não incluída na clusterização)</i>	-	<i>Mature</i>	<i>Emotionally immature, irresponsible</i>	3	<i>Wise, mature, polished</i>	<i>Dependent, silly, incoherent</i>	-
D/I	<i>Hypersensitive, infantile, sthenic emotionality/Sensitive, imaginative, anxious emotionality</i>	<i>Phlegmatic frustration tolerance/Rigid, tough, poise</i>	<i>self-deceiving (Não incluída)(Não incluída na clusterização)</i>	-	<i>Self-deceiving</i>	<i>Realistic</i>	5	<i>Neurotic</i>	<i>Not generally neurotic</i>	-

Na tabela acima, a relação entre cada traço de personalidade usado no modelo 12PF, a(s) variável(is) que pertence e seu(s) respectivo(s) fator(es) foi encontrada em CATTELL, 1946a, pp. 311-337).

Observou-se padrões similares aos identificados na comparação entre as variáveis do modelo de Cattell e os *clusters* obtidos pelo algoritmo SOM.

Mais especificamente observou-se esses padrões nas variáveis 1, 2, 6, 8, 9, 10, 11, 15, 19, 22, 31 e 32.

A variável 1 contém os traços *boastful* e *conceited*, que ficaram alocados pela rede SOM, respectivamente, nos *clusters* 74 e 75. Estes *clusters* são *clusters* vizinhos no grafo representando as ligações entre os traços selecionados.

A variável 2 do modelo 12PF contém os traços *clever* e *intelligent*, alocados pela rede nos *clusters* 3 e 18 respectivamente. Estes *clusters* também estão localizados em regiões próximas do mencionado grafo.

A variável 6 do modelo abriga os traços *grateful* e *friendly*, alocados respectivamente pela SOM no mesmo *cluster* (*cluster* 8). Ademais, o traço remanescente nesta variável (*hard*) foi alocado pela rede ao *cluster* 19, próximo ao *cluster* 8 no grafo.

Na variável 8, os traços *inflexible* e *severe* foram alocados aos *clusters* 32 e 33, vizinhos no grafo representando os relacionamentos entre os 222 traços selecionados para as aplicações de clusterização.

Similarmente, na variável 9 encontram-se os traços *grateful* e *hard*, alocados respectivamente nos *clusters* 8 e 19 pela rede neural. Estes clusters também se localizam em regiões próximas no grafo representando os relacionamentos entre estes traços no *thesaurus* adotado.

O mesmo ocorre em relação à variável 10, onde os traços *subjective* e *abstract*, presentes nesta variável, foram alocados em *clusters* próximos no referido grafo (*clusters* 85 e 86, definidos pela rede SOM).

Na variável 11 os traços *conscientious* e *painstaking* se encontram presentes. Estes traços foram alocados pela rede SOM no mesmo *cluster* (15); o mesmo se verificando na variável 15, onde os traços *genial* e *optimistic* foram alocados no mesmo *cluster* (8) pela rede neural. Nesta mesma variável, o traço *timid* foi alocado pela SOM no *cluster* 105, próximo ao *cluster* 8 no grafo.

A variável 19 é composta pelos traços *frank*, *cooperative* e *friendly*. Estes dois últimos traços foram alocados pela SOM no mesmo *cluster* (8) e o traço *Frank* foi alocado no *cluster* 1, próximo ao *cluster* 8 no grafo.

Na variável 22, os traços *social*, *responsive* e *genial* foram alocados no mesmo *cluster* pela SOM (8). O mesmo se verifica na variável 31, onde os traços *responsive*, *gregarious* e *sociable* foram alocados no mesmo *cluster* (8). Adicionalmente, o traço *hearty* foi alocado ao *cluster* 16, próximo ao *cluster* 8 no grafo representando as ligações entre estes traços no *thesaurus*.

Finalmente, na variável 32 se verifica que os traços *worried* e *anxious* foram alocados respectivamente aos *clusters* 97 e 117 pela rede neural, próximos no grafo.

Deste modo se verificou que em diversas variáveis (que por sua vez são *clusters* obtidos por Cattell, por meio das correlações entre as respostas de questionários contendo os referidos traços) do modelo 12 PF se observou que muitos traços nelas presentes se encontram nos mesmos *clusters* ou em regiões muito próximas do grafo representando o relacionamento entre os 222 traços selecionados, no *thesaurus*.

Isto revela que os *clusters* de traços no TTT usado não conseguiram reproduzir, as composições das variáveis desse modelo de personalidade (o mesmo se verificando em relação aos *clusters* obtidos pelo algoritmo MCL).

Similarmente ao verificado na comparação do modelo 12 PF com os resultados de clusterização obtidos pelo algoritmo MCL, somente 22,86% (8) das variáveis do modelo tiveram pelo menos 2 de seus traços (ou dos traços e elas correspondentes, no caso de traços

candidatos a substitutos de traços originalmente propostos por Cattell e ausentes no TTT) alocados em um mesmo *cluster* pela rede SOM.

Assim, 77,14% das variáveis do modelo de Cattell corresponderam a traços alocados a clusters diferentes pela rede SOM.

7.5 NOVA CONFIGURAÇÃO E TREINAMENTO DA REDE SOM

Frente aos resultados obtidos com redes SOM anteriormente e o fato da rede ter isolado muitos traços que originalmente não estavam isolados (isto é, que não possuíam ligação com nenhum outro) este trabalho procedeu a um novo treinamento de uma rede SOM, com outras configurações.

Neste esforço visou também aumentar o percentual de ligações *intra clusters* para os vértices de uma mesma componente conexa do grafo.

Deste modo, a primeira mudança implementada em relação às configurações anteriores foi se proceder à adoção de uma nova estratégia de codificação dos 222 traços considerados. Diferentemente do realizado anteriormente, se adotou a estratégia de se proceder a uma codificação independente para cada uma das duas maiores componentes conexas do grafo (componente 1, formada por 145 traços e componente 6, formada por 10 traços)

Assim, todas as demais componentes, por serem muito pequenas e por conterem no máximo três traços (a grande maioria contendo somente um traço), foram consideradas um *cluster* independente; não se justificando se realizar uma codificação seguida de um treinamento da rede SOM para cada uma delas.

Os vetores correspondentes à codificação em cada uma dessas duas componentes conexas se encontram no DVD anexo à esta tese.

Também os parâmetros da rede e a arquitetura do mapa topológico utilizado em cada uma dessas duas componentes conexas foi alterado. A seguir se apresenta estas configurações.

Configurações da rede SOM para a clusterização dos 145 traços na componente conexa 1 do grafo:

- **Arquitetura do mapa topológico utilizado:** Escolheu-se um mapa topológico para a rede neural de arquitetura hexagonal e com arranjo não aleatório (mais precisamente esse arranjo é composto de 08 linhas de hexágonos e 08 colunas de hexágonos).

- **Medida de distância utilizada:** Adotou-se como medida de distância entre os elementos sendo clusterizados, a distância Euclidiana, considerada a mais adequada para o problema e também se levando em consideração o pequeno porte do mesmo.

- **Taxa de aprendizagem utilizada:** Adotou-se como taxa de aprendizagem final para a rede o valor de 0,1.

- **Número de épocas:** se adotou 725 época.

- **Valor inicial para o tamanho das vizinhanças:** se adotou como valor inicial para este parâmetro o valor de 8.

- **Número de épocas para que o raio das vizinhanças vá para 1:** 700

- **Regra de atualização dos pesos sinápticos da rede neural:** utilizou-se a regra padrão (*default*) de atualização dos pesos sinápticos da *toolbox* de redes neurais do MATLAB.

Configurações da rede SOM para a clusterização dos 10 traços na componente conexa 6 do grafo:

- **Arquitetura do mapa topológico utilizado:** Escolheu-se um mapa topológico para a rede neural de arquitetura hexagonal e com arranjo não aleatório (mais precisamente esse arranjo é composto de 02 linhas de hexágonos e 02 colunas de hexágonos).

- **Medida de distância utilizada:** Adotou-se como medida de distância entre os elementos sendo clusterizados, a distância Euclidiana, considerada a mais adequada para o problema e também se levando em consideração o pequeno porte do mesmo.

- **Taxa de aprendizagem utilizada:** Adotou-se como taxa de aprendizagem final para a rede o valor de 0,1.

- **Número de épocas:** se adotou 50 épocas.

- **Valor inicial para o tamanho das vizinhanças:** se adotou como valor inicial para este parâmetro o valor de 2.

- **Número de épocas para que o raio das vizinhanças vá para 1:** 45

- **Regra de atualização dos pesos sinápticos da rede neural:** utilizou-se a regra padrão (*default*) de atualização dos pesos sinápticos da *toolbox* de redes neurais do MATLAB.

7.5.1 Resultados encontrados

A tabela 35 a seguir apresentará os resultados encontrados para a componente conexa 1 do grafo.

Tabela 35: Resultados encontrados na nova configuração SOM na componente conexa
1do grafo.

Número do cluster	Região do Mapa Topológico que recebeu palavras	Número da palavra	Traços	Componente Conexo do grafo
1	1	89	<i>strong</i>	1
1	1	92	<i>in shape</i>	1
1	1	93	<i>alcoholic</i>	1
1	1	94	<i>energetic</i>	1
1	1	95	<i>courageous</i>	1
1	1	96	<i>influential</i>	1
2	2	122	<i>honest</i>	1
3	4	62	<i>economical</i>	1
3	4	132	<i>frugal</i>	1
3	4	133	<i>provident</i>	1
3	4	134	<i>thrifty</i>	1
3	4	135	<i>sparing</i>	1
4	6	74	<i>gregarious</i>	1
4	6	84	<i>cosmopolitan</i>	1
4	6	85	<i>impartial</i>	1
5	7	79	<i>amiable</i>	1
5	7	121	<i>charming</i>	1
6	8	76	<i>genial</i>	1
6	8	86	<i>cheerful</i>	1
6	8	87	<i>resilient</i>	1
6	8	88	<i>optimistic</i>	1
7	9	6	<i>hearty</i>	1
7	9	90	<i>generous</i>	1
8	10	91	<i>devout</i>	1
8	10	123	<i>religious</i>	1
8	10	124	<i>pious</i>	1
8	10	125	<i>reverent</i>	1
9	13	73	<i>sociable</i>	1
9	13	82	<i>cooperative</i>	1
9	13	83	<i>coherent</i>	1
10	14	69	<i>social</i>	1
10	14	75	<i>convivial</i>	1
10	14	80	<i>affable</i>	1
10	14	128	<i>grateful</i>	1
11	15	77	<i>cordial</i>	1
12	16	117	<i>affectionate</i>	1

12	16	119	<i>responsive</i>	1
12	16	120	<i>kind-hearted</i>	1
13	17	8	<i>anxious</i>	1
13	17	140	<i>worried</i>	1
13	17	141	<i>panicky</i>	1
13	17	142	<i>afraid</i>	1
13	17	143	<i>frightened</i>	1
13	17	144	<i>timid</i>	1
13	17	145	<i>coward</i>	1
14	18	2	<i>enthusiastic</i>	1
14	18	7	<i>vivacious</i>	1
15	19	126	<i>frank</i>	1
15	19	127	<i>open-hearted</i>	1
16	20	21	<i>egotistical</i>	1
16	20	25	<i>opinionated</i>	1
16	20	26	<i>selfish</i>	1
17	22	81	<i>political</i>	1
18	23	65	<i>polished</i>	1
18	23	70	<i>informative</i>	1
18	23	71	<i>hospitable</i>	1
18	23	72	<i>poised</i>	1
19	24	78	<i>friendly</i>	1
19	24	118	<i>kind</i>	1
20	25	1	<i>excitable</i>	1
20	25	3	<i>irritable</i>	1
20	25	4	<i>high-strung</i>	1
20	25	109	<i>sour</i>	1
20	25	110	<i>sensitive</i>	1
21	27	18	<i>conceited</i>	1
21	27	19	<i>arrogant</i>	1
21	27	20	<i>boastful</i>	1
21	27	22	<i>self-satisfied</i>	1
21	27	23	<i>imitative</i>	1
21	27	24	<i>autocratic</i>	1
21	27	27	<i>self-confident</i>	1
22	29	57	<i>fresh</i>	1
22	29	116	<i>original</i>	1
23	30	58	<i>practical</i>	1
23	30	60	<i>sophisticated</i>	1
23	30	61	<i>constructive</i>	1
24	31	59	<i>discriminating</i>	1
24	31	63	<i>imaginative</i>	1
24	31	64	<i>aesthetic</i>	1
24	31	66	<i>eclectic</i>	1

24	31	130	<i>multifarious</i>	1
25	32	129	<i>many-sided</i>	1
25	32	131	<i>multiple</i>	1
26	33	5	<i>impulsive</i>	1
26	33	9	<i>instinctive</i>	1
26	33	12	<i>intuitive</i>	1
27	34	10	<i>emotional</i>	1
27	34	15	<i>sentimental</i>	1
28	35	16	<i>affected</i>	1
29	37	35	<i>thoughtful</i>	1
29	37	38	<i>observing</i>	1
30	38	41	<i>alert</i>	1
30	38	56	<i>active</i>	1
31	41	11	<i>mechanical</i>	1
31	41	13	<i>programmed</i>	1
31	41	14	<i>standardized</i>	1
32	42	17	<i>idealistic</i>	1
32	42	28	<i>visionary</i>	1
33	44	44	<i>rational</i>	1
33	44	67	<i>analytical</i>	1
33	44	68	<i>advisable</i>	1
34	45	40	<i>intelligent</i>	1
34	45	45	<i>logical</i>	1
34	45	46	<i>wise</i>	1
35	46	97	<i>severe</i>	1
35	46	100	<i>exact</i>	1
35	46	102	<i>ascetic</i>	1
35	46	106	<i>reclusive</i>	1
36	47	101	<i>ambitious</i>	1
36	47	104	<i>enterprising</i>	1
36	47	105	<i>industrious</i>	1
37	48	107	<i>reserved</i>	1
37	48	108	<i>inhibited</i>	1
38	49	37	<i>conjectural</i>	1
38	49	137	<i>doubtful</i>	1
38	49	138	<i>self-conscious</i>	1
39	50	34	<i>speculative</i>	1
39	50	36	<i>academic</i>	1
40	51	29	<i>ideal</i>	1
40	51	30	<i>abstract</i>	1
40	51	32	<i>conceptual</i>	1
40	51	33	<i>imaginary</i>	1
40	51	139	<i>subjective</i>	1
41	53	43	<i>clever</i>	1

41	53	49	<i>dexterous</i>	1
41	53	50	<i>witty</i>	1
42	54	42	<i>shrewd</i>	1
43	55	99	<i>inflexible</i>	1
44	56	103	<i>serious</i>	1
44	56	112	<i>dignified</i>	1
45	57	39	<i>mindful</i>	1
45	57	136	<i>cautious</i>	1
46	58	31	<i>hypothetical</i>	1
47	60	48	<i>skilful</i>	1
47	60	51	<i>deft</i>	1
47	60	52	<i>skilled</i>	1
47	60	53	<i>prepared</i>	1
47	60	54	<i>proficient</i>	1
47	60	55	<i>professional</i>	1
48	62	98	<i>austere</i>	1
48	62	111	<i>formal</i>	1
48	62	114	<i>conventional</i>	1
49	63	47	<i>hard</i>	1
50	64	113	<i>elevated</i>	1
50	64	115	<i>eloquent</i>	1

Como pode-se verificar na tabela 35 acima, com as escolhas de configurações utilizadas nota-se primeiramente que os clusters são formados por mais traços. A rede SOM, com estas configurações agora definiu 50 *clusters* a partir das 145 palavras contidas na componente conexa 1 do grafo.

A figura 34 abaixo apresenta a alocação desses 145 traços no mapa topológico da rede.

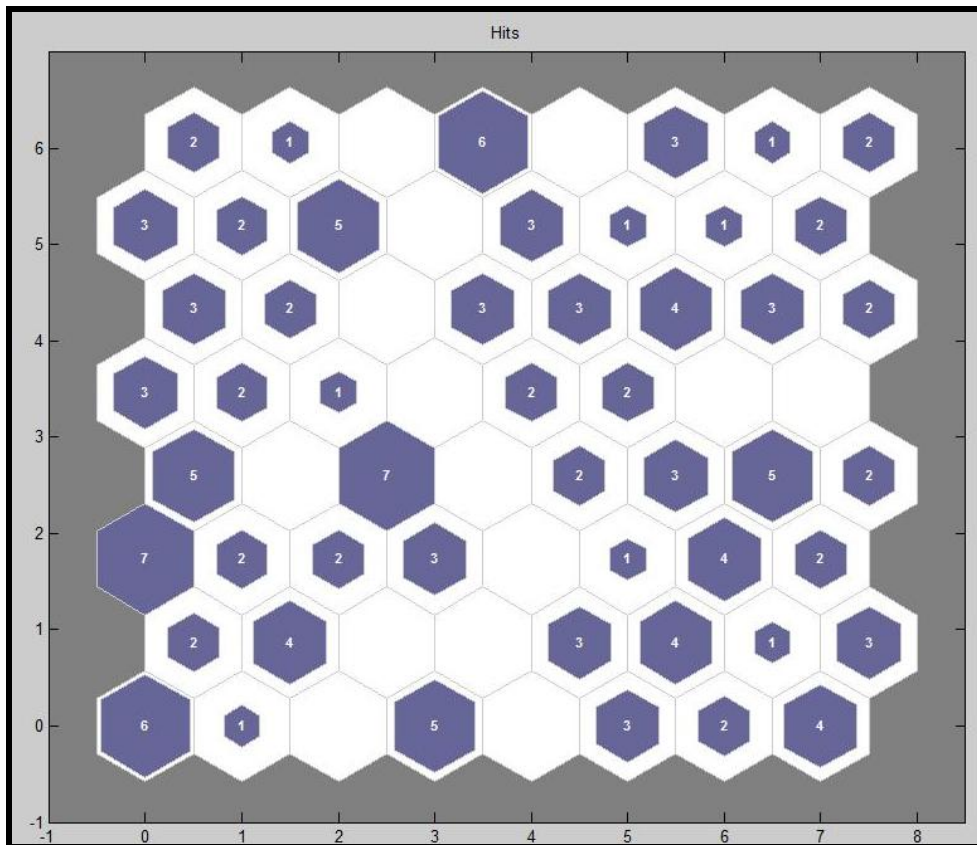


Figura 34: Alocação dos traços da 1ª componente conexa do grafo no mapa topológico

Fonte: Relatório *Sample Hits* da *Toolbox Neural Networks* do MATLAB.

A seguir se apresentará na tabela 36 abaixo os percentuais de ligações *inter* e *intra clusters* para os vértices desta componente conexa.

Tabela 36: Percentuais de ligações *intra* e *inter clusters* na 1ª componente conexa do grafo

<i>Clusters</i>	Número de ligações <i>intra cluster</i>	Número de ligações <i>inter clusters</i>	Total ligações	% de ligações <i>intra clusters</i>	% de ligações <i>inter clusters</i>	Número de traços no <i>cluster</i>
1	5	6	11	45,45%	54,55%	6
2	0	2	2	0,00%	100,00%	1
3	9	1	10	90,00%	10,00%	5
4	2	6	8	25,00%	75,00%	3
5	1	4	5	20,00%	80,00%	2
6	3	6	9	33,33%	66,67%	4
7	1	8	9	11,11%	88,89%	2
8	4	2	6	66,67%	33,33%	4
9	2	5	7	28,57%	71,43%	3

10	3	11	14	21,43%	78,57%	4
11	0	3	3	0,00%	100,00%	1
12	2	1	3	66,67%	33,33%	3
13	7	1	8	87,50%	12,50%	7
14	1	5	6	16,67%	83,33%	2
15	1	1	2	50,00%	50,00%	2
16	2	4	6	33,33%	66,67%	3
17	0	1	1	0,00%	100,00%	1
18	4	5	9	44,44%	55,56%	4
19	2	6	8	25,00%	75,00%	2
20	4	3	7	57,14%	42,86%	5
21	7	5	12	58,33%	41,67%	7
22	1	4	5	20,00%	80,00%	2
23	2	4	6	33,33%	66,67%	3
24	4	5	9	44,44%	55,56%	5
25	1	2	3	33,33%	66,67%	2
26	2	2	4	50,00%	50,00%	3
27	2	3	5	40,00%	60,00%	2
28	0	2	2	0,00%	100,00%	1
29	1	4	5	20,00%	80,00%	2
30	1	7	8	12,50%	87,50%	2
31	2	1	3	66,67%	33,33%	3
32	1	2	3	33,33%	66,67%	2
33	2	6	8	25,00%	75,00%	3
34	2	7	9	22,22%	77,78%	3
35	3	8	11	27,27%	72,73%	4
36	2	3	5	40,00%	60,00%	3
37	1	3	4	25,00%	75,00%	2
38	2	2	4	50,00%	50,00%	3
39	1	4	5	20,00%	80,00%	2
40	5	4	9	55,56%	44,44%	5
41	2	6	8	25,00%	75,00%	3
42	0	4	4	0,00%	100,00%	1
43	0	3	3	0,00%	100,00%	1
44	1	4	5	20,00%	80,00%	2
45	1	1	2	50,00%	50,00%	2
46	0	4	4	0,00%	100,00%	1
47	7	3	10	70,00%	30,00%	6
48	2	8	10	20,00%	80,00%	3
49	0	6	6	0,00%	100,00%	1
50	1	1	2	50,00%	50,00%	2

Como se pode notar, com as configurações adotadas os *clusters* encontrados pela rede agora possuem percentual substancialmente maior de ligações *intra clusters*. Alterações nessas configurações podem aumentar significativamente estes percentuais caso assim se deseje.

Finalmente, a figura 34 abaixo apresenta a U-matrix desta solução com redes SOM para esta componente conexa do grafo.

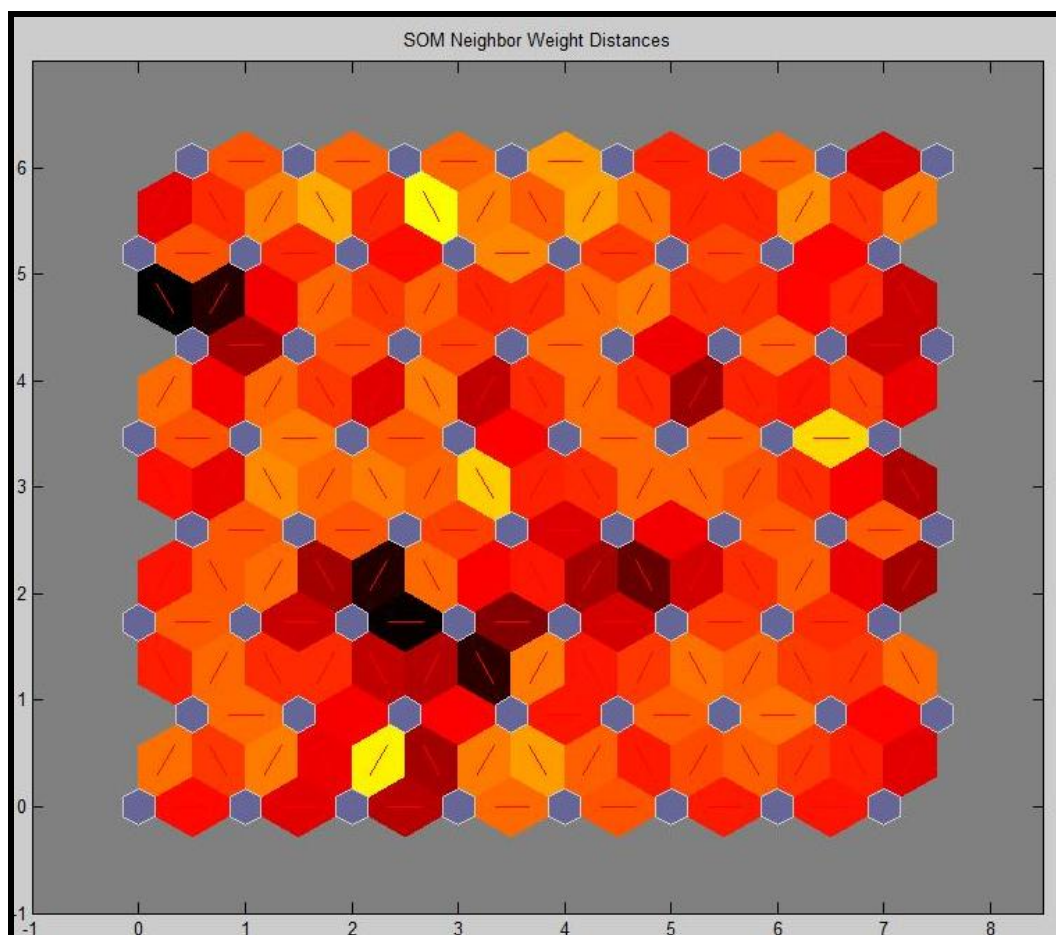


Figura 35: *U-Matrix* da nova clusterização dos traços na componente conexa 1 do grafo.

Fonte: Relatório *Neighbor weight Distances* da *Toolbox Neural Networks* do MATLAB.

Os resultados encontrados para a componente conexa conexa 6 do grafo foram idênticos aos encontrados pelo algoritmo MCL e por isso serão aqui omitidos.

Deste modo pode-se verificar que as características desejadas dos *clusters* podem ser alteradas mediante a mudança de seus parâmetros e da arquitetura do mapa topológico da rede. Em particular, pode-se buscar *clusters* tais que os vértices de uma determinada componente conexa do grafo possuam maior número de ligações *intra clusters* do que *inter clusters* mediante estes expedientes.

Outro ponto que se pode notar pelos resultados apresentados é que a mudança de estratégia de codificação dos vértices do grafo (passando-se a se codificar cada componente conexa independentemente) passou a apresentar número de vértices isolados em muito menor quantidade.

CAPÍTULO 8: APLICABILIDADES DAS ROTINAS COMPUTACIONAIS IMPLEMENTADAS EM ÁREAS RELACIONADAS ÀS TRAÇOS DE PERSONALIDADE

Neste capítulo se fará uma apresentação as aplicações que *clusters* de palavras podem ter nas áreas de modelos de personalidade, em aplicações organizacionais (sobretudo na avaliação psicológica e em processos seletivos, apresentando apenas sucintamente aplicações em outras áreas como Linguística, e *Knowledge Management*). Também pormenorizará como os desenvolvimentos computacionais desta tese podem auxiliar profissionais nesta área.

8.1 BREVE REVISÃO DAS MOTIVAÇÕES PARA O DESENVOLVIMENTO DAS ROTINAS COMPUTACIONAIS E DO SISTEMA PHP

Como anteriormente apresentado alguns dos principais estudos na área de traços de personalidade foram estudos psicoléxicos; estudos que se valiam da hipótese lexicográfica (que, como visto, enuncia que as diferenças individuais mais socialmente relevantes estão codificadas na linguagem natural), hipótese esta que continua a ser amplamente adotada e aceita pela maioria dos pesquisadores na área.

Foi verificado anteriormente neste trabalho que para alguns autores pesquisadores da área de TP (com, por exemplo, JACKSON e PAUNONEN, 2000) a pesquisa nesta área ainda está aberta a contribuições e pesquisas, especialmente porque as evidências empíricas têm evidenciado de que os modelos de cinco fatores, os mais aceitos atualmente na área, não tem conseguido explicar completamente as diferenças individuais ao longo das diversas culturas e que o número de fatores necessários para tanto varia em algumas regiões do globo. JACKSON e PAUNONEN, 2000, que mostram que a própria organização (e não sooo número) desses fatores varia de cultura em cultura. Mesmo a versão mais aceita dentre os modelos de cinco fatores (apresentada por GOLDBERG, 1993) tem sido criticado a esse respeito.

Mesmo não se variando a cultura, segundo estes mesmos autores, não há consenso sobre o número de fatores capazes de descrever a personalidade, nestes modelos, assim como das variáveis que o compõe, de forma que este campo ainda está aberto a muitas pesquisas (JACKSON e PAUNONEN, 2000). Um exemplo desta situação é apresentado no

estudo de SINGH e MISRA (2010) onde cinco fatores não foram suficientes para explicar as variações de personalidade; sendo necessária a inclusão de mais fatores.

Retornando aos estudos psicoléxicos anteriormente feitos, notou-se também que diversos deles utilizaram-se da coleta de termos de um dicionário (físico) efetuada pelo estudo de ALLPORT e ODBERT (1936), como exposto no capítulo.

Esses estudos se valeram, como etapa inicial, de algum método de organização e classificação dos traços utilizados. No estudo de ALLPORT e ODBERT (1936), a coleta foi feita de maneira manual e a classificação dos traços foi feita de forma exaustiva pela leitura de um a um, seguida da alocação de cada um desses traços em quatro grandes categorias pré-definidas. Estudos posteriores (como o de CATTELL, 1943a e b) também se valeram de esquemas de classificação de traços em categorias (de acordo com suas similaridades quanto aos comportamentos a que se referiam). CATTELL (1943a e b), por exemplo, a partir de uma seleção dos termos coletados por ALLPORT e ODBERT (1936), em um primeiro momento os organizou em categorias de sinônimo e posteriormente selecionou 171 desses termos para posteriores estudos que culminaram com o modelo de 16 fatores da personalidade.

Por outro lado, a linguagem é uma das melhores expressões de uma determinada cultura, na medida em que reflete diretamente elementos, conceitos e valores de uma determinada sociedade. Esta linguagem é dinâmica na medida em que os significados atribuídos às palavras podem mudar ao longo do tempo (e desta forma também os termos utilizados para descrever determinados comportamentos também podem mudar). Os dicionários, por sua vez (e os bancos de dados do tipo *thesaurus*), são uma forma efetiva de se acessar como uma determinada cultura atribui sentidos através das palavras; refletindo por tabela, que palavras ela utiliza para representar os diferentes comportamentos humanos.

Deste modo, como a área de traços ainda apresenta uma série de desafios, pesquisas que visem superar as limitações atuais, como por exemplo, as mencionadas com os modelos de cinco fatores são bem vindas. Estas pesquisas podem incluir, por exemplo, a construção de modelos de personalidade que se atualizem quanto às possíveis mudanças que possam surgir nos sentidos atribuídos a determinados termos - e inclusive para darem conta do possível surgimento de novas palavras para designarem determinados comportamentos humanos - e que visem superar as dificuldades anteriormente verificadas.

Nesta linha de raciocínio, e levando-se em consideração as propriedades de CPT já enunciadas anteriormente, esta tese vislumbrou e implementou um conjunto de soluções

computacionais que pudessem fornecer ao pesquisador (e outros profissionais da área) métodos sistemáticos de organização desses traços em um *thesaurus*, visando com isso por um lado agilizar as etapas iniciais de um estudo psicológico (referentes à coleta e organização de traços de personalidade em categorias de similaridade) assim como fornecer outros subsídios às atividades desses profissionais, como será visto mais a frente.

Como os CPT's reúnem traços de acordo com suas proximidades semânticas tal como presente na língua natural esses agrupamentos representam uma forma útil de organização desses traços para as etapas posteriores de um estudo psicológico.

Essa possibilidade decorre do fato de que, como mencionado anteriormente, existem muitas palavras descritoras da personalidade (muitas delas possuem significados similares) o que implica, na prática, em certo grau de sobreposição dos TP's quanto aos seus significados comportamentais. Isto impõe necessariamente, ao pesquisador, a necessidade de se agrupar tais traços em função de suas similaridades quanto aos aspectos comportamentais que representam e a partir daí permitir ao pesquisador:

- 1) eliminar termos muito redundantes e a escolha dos termos que representam de forma mais abrangente e exaustiva possível as diversas manifestações da personalidade (no maior número possível de sociedades e culturas), de forma a se chegar a um conjunto "enxuto" de traços; e,

- 2) definir as dimensões da personalidade relevantes em função das correlações de respostas em questionários envolvendo tal conjunto de traços e da aplicação de técnicas estatísticas que permitam reduzir ainda mais esse conjunto de traços (variáveis) em fatores com a menor perda de informação possível.

Deste modo, e também como já ressaltado anteriormente, os CTP's podem auxiliar o pesquisador na escolha inicial dos traços que descrevam da forma mais abrangente possível os aspectos comportamentais presentes nos *clusters* a que pertencem; uma vez que agrupa tais traços de acordo com suas proximidades semânticas segundo a língua natural.

8.2 BREVE REVISÃO DOS RELATÓRIOS GERADOS PELAS ROTINAS E PELO SISTEMA DESENVOLVIDOS

Inicialmente deve-se notar que por meio das diversas formas de representação do grafo subjacente ao TTT utilizado (ou de um subconjunto de palavras deste) podem-se obter informações de como os traços estão relacionados entre si tanto dentro de um *cluster*, como entre *clusters* diferentes.

Estas diversas formas de representação (como matrizes de adjacência e listas de adjacência) podem auxiliar na determinação não só se duas palavras quaisquer, consideradas para uma determinada aplicação, possuem associação entre elas como a que *clusters* pertencem.

Estas informações, além de possuírem grande importância em diversas áreas do conhecimento (como será a seguir apresentado), tornam-se de difícil obtenção conforme a massa de dados cresce e particularmente quando os pesquisadores ou usuários não dispõem de habilidades específicas (em particular relativas a processamento de dados, entendimento de complexidade algorítmica etc.) e não podem dispor de uma equipe de trabalho que possa lhes ajudar.

A esse respeito, o desenvolvimento do sistema em linguagem PHP, contando com interfaces gráficas cujas utilizações são de fácil aprendizado permite aos usuários gerar tais relatórios de forma fácil, sem necessidade de se desenvolver nenhum código. Neste sistema os usuários são requisitados basicamente a estabelecerem o caminho do arquivo de origem e a partir daí pressionarem as teclas de comando e configurações das rotinas, como anteriormente apresentado.

8.2.1 Observações gerais sobre os desenvolvimentos

Como não existe um consenso entre pesquisadores e profissionais da área no que diz respeito aos traços a serem incluídos em um modelo que vise explicar a personalidade humana; as rotinas implementadas nesta tese permitem que o usuário facilmente retire ou adicione traços ao seu estudo, gerando relatórios com o conjunto de palavras de interesse.

O fato de que *clusters* de palavras neste tipo de base de dados agruparem termos segundo sentidos/conceitos mais amplos, englobando os termos em cada agrupamento deste tipo e de que palavras vizinhas cumprem na maioria das vezes função de sinônimas entre si pode fornecer *insights* para o pesquisador/profissional sobre quais palavras devem ser incluídas nas suas aplicações e que palavras representam de forma mais abrangente e exaustiva as ideias de cada *cluster*.

Quando se pode construir tais relatórios com facilidade e relativa eficiência computacional obtém-se um ganho de produtividade.

Neste sentido, os programas e códigos desenvolvidos contribuem em primeiro lugar, para a facilitação da pesquisa nas áreas relacionadas a TP, em particular para os usuários que não dispõem das capacitações em desenvolvimento, programação, banco de dados etc.

O pesquisador/profissional poderá decidir também que relatórios serão gerados a respeito do grafo associado à rede de ligações entre os traços selecionados, no *thesaurus*. Como uma determinada aplicação da teoria de traços pode se preocupar com um pequeno conjunto de traços específicos (como por exemplo, em um processo seletivo para um dado posto de trabalho, onde as habilidades e características comportamentais ideais para o posto são identificadas a priori e descritas por um conjunto de TP's) a possibilidade do pesquisador/profissional ter a liberdade de incluir somente os traços que lhe convier passa a representar grande valor para as áreas do conhecimento relacionadas a estas questões. Como visto, as ferramentas desenvolvidas ao longo deste trabalho propiciam essa flexibilidade.

As rotinas implementadas também representam ferramentas de alta valia do ponto de vista gerencial e organizacional, pois, como será visto mais adiante, além da possibilidade do profissional da área competente (por exemplo, da área de recursos humanos) incluir os traços (desde que existam no TTT) que lhe convier em uma dada aplicação; ainda fornecem subsídios para a avaliação de indivíduos (ou grupos de trabalho) de forma integrada com os demais relatórios produzidos pelas rotinas utilizadas.

Pelo fato das rotinas implementadas e dos programas desenvolvidos relativos ao preparo de relatórios referentes à estrutura dos grafos citados (assim como do algoritmo MCL de clusterização) terem sido elaborados (e otimizados) em linguagem Java, o pesquisador/profissional passa a contar com o recurso de portabilidade desta linguagem; o que significa na prática, que os programas desenvolvidos rodam em qualquer arquitetura de *hardware* e em qualquer sistema operacional; desde que a *Java Virtual Machine* (JVM) para o sistema operacional utilizado esteja instalada a priori. Isso significa que, havendo necessidade ou conveniência o pesquisador/profissional pode rodar e gerar os relatórios mencionados em qualquer computador que disponha do JVM (não ficando preso às limitações de um determinado hardware por conta da impossibilidade de migrar para outro sistema operacional).

Ambos os métodos de clusterização empregados possuem parâmetros capazes de alterar as características dos *clusters* finais encontrados; de forma que estas metodologias permitem um ajustamento fino dos *clusters* finais às exigências adicionais que possam ser impostas aos mesmos para satisfazerem um determinado problema de clusterização em particular (a esse respeito, vide, por exemplo, as condições adicionais exigidas nas aplicações envolvendo os 222 TP nesta tese).

8.2.1.1 Considerações gerais sobre as atratividades das soluções computacionais implementadas

Em linhas gerais as soluções computacionais implementadas neste trabalho tem como principais atrativos para os diferentes públicos potenciais os seguintes aspectos:

- Implementam algoritmos otimizados em linguagem baixa (no caso dos códigos em linguagem Java),
- Propiciam aos usuários a possibilidade de gerar diversos relatórios relacionados a Teoria de Grafos.

Com aplicações deste tipo (isto é, envolvendo palavras do *thesaurus* empregado) este recurso possibilita aos usuários envolvidos:

- Selecionarem as palavras do TTT com as quais desejam desenvolver análises e gerar relatórios.
- Adicionalmente, pela própria natureza da planilha Excel (formato escolhido para efetuação de importação de palavras no banco de dados), na ausência de uma palavra de interesse no TTT (ou mesmo na eventual discordância quanto ao relacionamento entre duas palavras quaisquer na planilha), pode-se incluir novas palavras no mesmo, ou mesmo se alterar os relacionamentos entre palavras; bastando, para isso, a mudança de célula (no Excel) de uma determinada palavra de uma célula da planilha para outra; este fato também dá aos usuários potenciais flexibilidade de uso das ferramentas.
- Finalmente, quase todos os relatórios produzidos pelo sistema Java são arquivos em texto; de forma que os mesmos podem ser manualmente alterados ou customizados pelo usuário, caso haja necessidade. Este recurso é particularmente atrativo, por exemplo, em uma situação onde o usuário não concorde com a presença de determinada palavra em um dado *cluster* e deseje mudá-la para outro; ou ainda caso este mesmo usuário deseje atribuir um determinado nome a um *cluster* - refletindo as ideias gerais das palavras contidas no mesmo -.

Assim, por exemplo, se no relatório de *clusters* (um dos relatórios relacionados ao algoritmo MCL), produzido a partir de uma importação de palavras do *thesaurus*, o usuário não concordar com a presença de uma dada palavra, ele poderá deslocá-la para outro *cluster* onde ache sua presença mais propícia; dependendo da aplicação com que se depara.

Como se verá a seguir, diversas são as áreas do conhecimento onde o problema de se identificar relacionamentos semânticos entre palavras é reconhecidamente importante e vem recebendo contribuições.

8.3 APLICABILIDADES DAS ROTINAS E SISTEMA DESENVOLVIDOS

Na literatura científica, o problema de se encontrar *clusters* em grafos é bastante conhecido e tratado; sobretudo nas áreas de Processamento de Dados (por exemplo), com a escolha de arquiteturas ótimas de processamento paralelo em redes de computadores, CHRETIENNE, 1989, SARKAR, 1989, GERASOULIS E YANG, 1992 e CHALUPA E POSPICHAL, 2012, internet (com o desenvolvimento de mecanismos de buscas na web baseados em *clusters* de palavras como em BRASLAVSKI et al, 2004), *Knowledge Management* (particularmente em classificação de documentos, A esse respeito vide PARVIN et al, 2012), além de aplicações na Psicologia, como será apresentado a seguir.

A seguir se apresentará possíveis aplicações das rotinas implementadas e desenvolvidas neste trabalho.

8.3.1 Aplicações psicológicas

8.3.1.1 Aplicações psicológico-organizacionais

Dentro das organizações humanas (com ou sem fins lucrativos, privadas ou públicas, civis ou militares etc.) a avaliação psicológica de indivíduos se constitui um importante tema.

Como ressalta DE RAAD (2000, p.4), recentemente se verificou um crescimento do interesse e do uso dos modelos de personalidade e taxonomia envolvida nos mesmos em aplicações cínicas e em processos seletivos organizacionais. Nas palavras do próprio autor:

“Recently, interest in the Five-Factor taxonomy boomed enormously. Notwithstanding this growing popularity of the Big Five Model and the emergence of its utility in, for example, clinical and health psychology ... and personnel selection...., considerably more research is needed to enhance the precise meaning of the factors, to improve their validity, and to arrive at their cross-cultural applicability”. (DE RAAD, 2000, p.4).

Ainda, como corrobora CATTELL e MEAD (2007), a respeito das aplicações que o Questionário de 16 Fatores de Personalidade recebeu ao longo dos anos:

“Due to its scientific origins, the 16PF Questionnaire has a long history of empirical research and is embedded in a well-established theory of individual differences....A conservative estimate of 16PF research since 1974 includes more than 2,000 publications (Hofer and Eber, 2002). Most studies have found the 16PF to be among the top five most commonly used normal-range instruments in both research and practice (Butcher and Rouse, 1996; Piotrowski and Zalewski, 1993; Watkins et al., 1995). The measure is also widely used internationally, and since its inception has been adapted into over 35 languages worldwide”. (CATTELL e MEAD, 2007, p. 135).

Estas avaliações visam principalmente a acompanhar o desenvolvimento pessoal de cada membro de uma determinada organização (como por exemplo, a sua evolução em termos de suas características de personalidade, tais como: capacidade de lidar com diferentes ambientes de trabalho, níveis de pressão, maturidade emocional etc.).

Outro papel do processo de avaliação psicológica e de *personality assessment* nas organizações é avaliar a adequabilidade de um potencial candidato (externo à organização) a um posto de trabalho ou a adequabilidade de um funcionário a uma vaga existente (no caso de um processo de recrutamento interno).

Deste modo, frequentemente estas organizações põem em prática métodos de avaliação nesse sentido.

Assim, o sistema em PHP e as rotinas implementadas em Java podem ser utilizados por estas organizações, para auxiliá-las em tais propósitos. Deste modo, as funcionalidades relacionadas à determinação de *clusters* de traços, conjuntamente com a possibilidade de se atribuir *ratings* aos mesmos por meios de questionários ficam disponíveis para o auxílio na execução de tais processos.

A seguir se detalhará como os desenvolvimentos computacionais desta tese podem auxiliar neste sentido.

8.3.1.1.1 Recrutamento e seleção

Em processos de recrutamento e seleção características de candidatos a postos de trabalho são levantadas mediante métodos de avaliação psicológica. Assim, questionários que atribuam *ratings* a traços de interesse podem ser empregados e as ferramentas de clusterização de traços e de elaboração de relatórios de *clusters* podem ser imediatamente usadas para auxiliar o desenvolvimento desses relatórios.

Nos casos de avaliações periódicas de funcionários de uma determinada organização ou em processos de seleção interna, estas avaliações podem ser feitas, pelo Gerente do funcionário sendo avaliado ou por outros funcionários que conheçam bem (ou trabalhem com) este funcionário; e podem utilizar um conjunto de traços de interesse provindos do *thesaurus* (ou adicionados ao mesmo, caso não constem inicialmente no seu acervo de palavras).

No caso de um processo seletivo externo, estas avaliações podem ser feitas mediante as observações, por parte dos psicólogos encarregados do processo seletivo, de cada um dos candidatos participantes do processo, ao longo de suas etapas (incluindo as entrevistas e dinâmicas).

Ao obter *clusters* de traços agrupados em função do grau de similaridade entre si, os profissionais envolvidos em tais processos (tais como profissionais de Recursos Humanos, psicólogos organizacionais e/ou gerentes participantes dos mesmos) podem, adicionalmente aos *ratings* dados nas respostas aos questionários envolvendo tais traços, atribuírem pesos diferenciados a cada um dos CTP's, dependendo de que características comportamentais sejam mais desejáveis para o posto em questão. Com estes procedimentos podem atribuir uma nota final a cada candidato.

Com respeito a estas possibilidades algumas das rotinas desenvolvidas em linguagem Java permitem a seleção inicial de traços de interesse no *thesaurus*, a geração de relatórios onde os *clusters* de traços identificados pelo algoritmo MCL são listados (no relatório de *clusters*), a geração de questionários contendo estes traços do relatório de *clusters* ou parte deles, em planilha Excel.

As rotinas permitem ainda que os questionários preenchidos (contendo os *ratings* de cada candidato quanto a cada traço selecionado) sejam utilizados para cada traço envolvido em conjunto com os pesos dados a cada *cluster*, pelo analista).

Deste modo, resumidamente, por meio das rotinas implementadas em Java, o profissional encarregado do processo seletivo pode:

- Selecionar os traços de personalidade de interesse em um determinado processo seletivo, presentes no *thesaurus*.
- Identificar agrupamentos de traços segundo a similaridade entre eles (no que se refere aos diferentes aspectos da personalidade de acordo com o algoritmo MCL). Pode ainda customizar este relatório, dando nomes a cada *cluster* identificado por este algoritmo (por exemplo, atribuindo um nome que represente os aspectos gerais de personalidade de cada agrupamento encontrado).
- Visualizar como os traços escolhidos e os *clusters* de traços encontrados estão relacionados, segundo a linguagem comum, no *thesaurus* (usando o *software R*).
- Selecionar (caso ache relevante) um conjunto mais restrito (menor) de traços de interesse, a partir da análise dos CTP's definidos pelo MCL (especialmente porque os traços assim dispostos podem ajudar o profissional a melhor entender em que medida traços pertencentes a um mesmo *cluster* representam comportamentos similares e/ou mesmo idênticos; de modo que o profissional possa escolher aqueles termos que representem de forma mais abrangente os comportamentos de interesse, em cada um desses agrupamentos).
- Gerar questionários (em planilha Excel) a partir dos traços selecionados.

- Recolher o questionário preenchido e atribuir pesos diferenciados (se for o caso) a cada *cluster* de traços de forma a privilegiar os *clusters* com as características mais importantes para o exercício de determinada função.

- Gerar um relatório de *ratings* do candidato.

Estas possibilidades apresentadas pelo conjunto de rotinas (em linguagem Java) possibilitam grande flexibilidade de trabalho a estes profissionais; utilizando os traços presentes no *thesaurus* (ou incluídos manualmente neste banco pelo próprio profissional) como insumo para elaboração desses relatórios e para o processo de seleção.

Alternativamente, muitos desses processos de seleção (interna e externa) e avaliação empregam redações elaboradas pelos próprios candidatos. Estas redações podem ter como tema uma auto-descrição do candidato, relatando, por exemplo, por meio de experiências pessoais e profissionais, suas características pessoais.

Nestas redações, traços importantes podem ser identificados de forma que a avaliação de um indivíduo possa se dar em função dos *ratings* por ele obtido em cada um desses traços e dos pesos atribuídos aos mesmos, (assim como dos pesos atribuídos aos *clusters* a que pertençam estes traços relevantes para o processo).

Textos contendo informações importantes que possam auxiliar de redações/anotações feitas pelos seus avaliadores ao longo de um determinado período de tempo. Estas anotações trazem as percepções do avaliador em um processo seletivo (ou, por exemplo, de um gerente, no caso de avaliações periódicas de funcionários já contratados ou em casos de processos seletivos internos) a cerca dos indivíduos sendo avaliados. Tais relatórios, também podem ser lidos pelas rotinas mencionadas, com a produção dos mesmos resultados já mencionados, desde que estejam nos formatos requeridos por estas rotinas.

Assim, algumas das rotinas desenvolvidas em Java neste trabalho permitem ao usuário, a partir de um conjunto de *clusters* de traços e da leitura dos mencionados textos o profissional a cargo de tal processo pode rapidamente, por meio das rotinas elaboradas:

- Identificar que traços de interesse estão presentes nestes textos/redações contendo as avaliações dos candidatos.

- Identificar a frequência com que estes traços estão presentes.

- Associar cada traço de interesse presente nestes textos (assim como sua frequência) com os *clusters* a que pertencem.

- Associar os *ratings* obtidos por cada candidato (por meio da aplicação de um questionário) a cada traço.

- Atribuir (ou não) pesos diferenciados a cada *cluster* e computar *ratings* finais para cada candidato, que representariam uma medida numérica do grau de adequabilidade do candidato ao posto em questão, levando-se em consideração somente suas características a, o conjunto de traços selecionados para avaliação, os *ratings* obtidos por cada candidato nos questionários e os pesos dados a cada *cluster*.

Como se pode depreender do que foi anteriormente exposto, os textos contendo avaliações do candidato podem ter diferentes origens:

- Podem ser obtidos de uma redação, elaborada pelo próprio candidato e contendo respostas a perguntas selecionadas (como por exemplo, perguntas que permitem ao candidato apresentar seus pontos fortes e fracos, tanto do ponto de vista técnico/profissional como em termos de características pessoais para o cargo considerado) ou relatos pessoais de suas experiências passadas.

- Relatórios elaborados pela equipe encarregada do processo seletivo, contendo observações sobre cada candidato ao longo do processo e das dinâmicas do mesmo. O exemplo a seguir ilustra um exemplo de como as rotinas desenvolvidas e os relatórios gerados com elas podem auxiliar tais profissionais.

Assim, rotinas que identifiquem a presença de traços de interesse e permitam a geração de relatórios a serem utilizados na avaliação psicológica de indivíduos em contextos organizacionais (relatórios estes que podem ainda ser customizados, dependendo das necessidades dos usuários) fornecem ao profissional flexibilidade de trabalho.

8.3.1.1.1 Exemplo de utilização das rotinas desenvolvidas em uma aplicação organizacional (fictícia) envolvendo um processo seletivo.

Para fins de ilustração de como as rotinas desenvolvidas neste trabalho podem auxiliar diversas organizações em processos seletivos, se irá supor aqui que uma empresa, por meio do seu departamento de Recursos Humanos, identificou uma vaga na área de atendimento a clientes preferenciais, para recrutamento interno.

Os psicólogos encarregados do processo selecionaram um conjunto inicial de traços, correspondentes aos 222 apresentados no capítulo 7. Procederam então à clusterização destes dados, por meio do algoritmo MCL, com o propósito de obterem informação de quão similares estes traços são, utilizando como critério, a linguagem natural (tal como retratada no *thesaurus* adotado nesta tese). O Relatório de *Clusters*, produzido por uma das rotinas em Java, para este conjunto de traços é apresentado abaixo:

Tabela 37: Relatório de *clusters* do algoritmo MCL gerado por rotina em Java

Cluster 1: intelligent, advisable, alert, logical, shrewd, witty, proficient, clever, skilled, wise, dexterous, professional, active, prepared, skilful, analytical, rational, observing, deft.

Cluster 2: discriminating, imaginative, multifarious, eclectic, many-sided, multiple.

Cluster 3: cosmopolitan, social, cordial, hospitable, cheerful, gregarious, poised, kind, aesthetic, coherent, kind-hearted, friendly, affable, affectionate, charming, polished, sociable, grateful, political, amiable, convivial, impartial, genial, cooperative, responsive, informative.

Cluster 4: programmed, mechanical, standardized, instinctive, intuitive.

Cluster 5: impulsive, hearty, enthusiastic, resilient, optimistic, generous, vivacious.

Cluster 6: sentimental, idealistic, emotional.

Cluster 7: imaginary, hypothetical, ideal, subjective, conceptual, abstract, visionary.

Cluster 8: sour, sensitive, inhibited, reclusive, high-strung, excitable, reserved, irritable.

Cluster 9: open-hearted, frank, devout, honest, pious, religious, reverent.

Cluster 10: constructive, sophisticated, practical.

Cluster 11: energetic, in shape, influential, original, strong, alcoholic, courageous, fresh.

Cluster 12: selfish, arrogant, self-satisfied, self-confident, opinionated, imitative, boastful, egotistical, autocratic, affected, conceited.

Cluster 13: mindful, conjectural, academic, doubtful, self-conscious, thoughtful, cautious, speculative.

Cluster 14: serious, ascetic, ambitious, industrious, hard, inflexible, dignified, exact, austere, formal, conventional, eloquent, elevated, severe, enterprising.

Cluster 15: afraid, worried, frightened, panicky, coward, timid, anxious.

Cluster 16: frugal, thrifty, sparing, provident, economical.

Cluster 17: fancy.

Cluster 18: fastidious.

Cluster 19: flattering.

Cluster 20: preventive.

Cluster 21: corporal, corporeal, bodily, physical, animal.

Cluster 22: leading, interested, visible, curious, perceivable, prominent.

Cluster 23: acquisitive.

Cluster 24: amorous.

Cluster 25: pugnacious, argumentative.

Cluster 26: speaking, assertive, vocal, unwritten.

Cluster 27: athletic.

Cluster 28: closed.

Cluster 29: complaining, sarcastic, cynical.

Cluster 30: tactful, painstaking, conscientious.

Cluster 31: contented.

Cluster 32: persevering, patient, controlled.

Cluster 33: musical, decorative.

Cluster 34: easygoing.

Cluster 35: eccentric.

Cluster 36: effeminate.

Cluster 37: evasive.

Cluster 38: gluttonous.

Cluster 39: governing.

Cluster 40: habituated.

Cluster 41: mulish, headstrong.

Cluster 42: hurried.

Cluster 43: independent.

Cluster 44: influencing.

Cluster 45: inhuman.

Cluster 46: introspective.

Cluster 47: jealous.

Cluster 48: laughing.

Cluster 49: leader.

Cluster 50: verbal, lexical.

Cluster 51: mischievous.

Cluster 52: mystical.

Cluster 53: neurotic.

Cluster 54: observer.

Cluster 55: pedantic.

Cluster 56: perfidious.

Cluster 57: perfunctory.

Cluster 58: pictorial.

Cluster 59: pitying.

Cluster 60: plaintive.

Cluster 61: reliable, stable.

Cluster 62: slanderous.

Cluster 63: slow.

Cluster 64: splenetic.

Cluster 65: sportive.

Cluster 66: talkative.

Cluster 67: temperate.

Cluster 68: vindictive.

Cluster 69: visual.

Cluster 70: wandering.

O relatório apresentado acima é escrito em formato .txt separado por tabulações. Não contém originalmente a moldura apresentada na tabela. Este relatório foi gerado a partir do conjunto de 222 traços selecionados a partir dos 171 originalmente propostos por CATTELL (1943b) e é escrito pela rotina de forma a incluir somente os *clusters* formados por mais de um traço, pelo algoritmo MCL; devendo ser alterado manualmente para incluir os demais *clusters* contendo traços que originalmente já estavam isolados na seleção de termos feita no *thesaurus*.

Supondo-se que após análise desse relatório os profissionais responsáveis pelo processo seletivo tenham chegado à conclusão que os traços de personalidade mais importantes e relevantes para o cargo se encontravam reunidos nos *clusters* 1, 2, 3, 11, 26, 32 e 50.

Nestes *clusters*, se irá assumir que, estes profissionais tenham procedido à seleção dos traços que representavam, da forma mais abrangente possível os comportamentos presentes em cada um deles.

Deste modo, a seleção final de traços de maior interesse para utilização na avaliação psicológica dos candidatos ao processo seletivo em questão ficou contendo os seguintes traços: *analytical (cluster 1)*, *eclectic (cluster 2)*, *sociable (cluster 3)*, *assertive (cluster 26)* e *patient (cluster 32)*.

Com base nestes traços, geraram, com as rotinas aqui implementadas, um questionário em Excel, para ser preenchido pelos avaliadores de cada candidato. Os avaliadores desses candidatos se restringiram aos seus gerentes imediatos, que convivem diariamente com os mesmos.

Estes questionários, depois de preenchidos na própria planilha Excel, deveriam ser devolvidos ao setor de RH, por e-mail.

A tabela abaixo apresenta um exemplo de questionário preenchido por um gerente responsável por um dos candidatos a vaga.

Tabela 38: Exemplo de questionário gerado a partir do relatório de *clusters*, por rotina em Java e preenchido para fins de ilustração (contendo somente os traços selecionados e seus antônimos).

<i>Cluster</i>	Traço	<i>Rating</i>	Antônimo	Peso
<i>Cluster 01</i>	<i>analytical</i>	5	<i>non-analytical</i>	1
<i>Cluster 02</i>	<i>eclectic</i>	4	<i>non-eclectic</i>	1
<i>Cluster 03</i>	<i>sociable</i>	3	<i>non-sociable</i>	2
<i>Cluster 26</i>	<i>assertive</i>	2	<i>non-assertive</i>	1
<i>Cluster 32</i>	<i>patient</i>	4	<i>non-patient</i>	1

O relatório acima é criado em formato. xls. Não inclui as bordas apresentadas na tabela.

É escrito inicialmente com pesos 1 para cada *cluster* e o usuário (alguém que irá avaliar uma terceira pessoa por meio deste relatório) deve preencher, ao longo da coluna intitulada “*Rating*”, um valor numérico entre 0 e 5. Valores próximos a cinco significam que a pessoa sendo avaliada apresenta fortemente a característica descrita pelo traço na coluna “Palavra”. Valores próximos a 0 implicam que a pessoa possui características próximas aos conceitos opostos aos dos traços listados ao longo desta coluna (estes conceitos opostos são listados na coluna “Antônimo” do questionário).

Finalmente, o usuário pode alterar manualmente, na planilha gerada pelo Excel (ou por meio da rotina), os pesos dados a cada *cluster*. Os valores apresentados no questionário acima foram preenchidos para fins ilustrativos.

Encaminhando esses questionários aos gerentes diretamente responsáveis pelos candidatos à vaga, puderam receber *ratings* desses candidatos de pessoas que os conheciam razoavelmente bem no ambiente profissional.

A tabela 38 acima apresenta um exemplo de um questionário produzido pelas rotinas Java a partir dos traços selecionados, preenchido. Com base nesses *ratings*, os responsáveis pelo processo seletivo têm mais um critério de avaliação dos candidatos à sua disposição.

8.3.1.1.1.2 Segundo exemplo de utilização das rotinas desenvolvidas em uma aplicação organizacional (fictícia) envolvendo um processo seletivo.

Como já mencionado anteriormente, outros instrumentos que proveem insumo para avaliações psicológicas são documentos que contenham discursos ou entrevistas manuscritas

da pessoa sendo avaliada; ainda ou textos/redações produzidas por alguém que conheça bem esse indivíduo ou produzidas por ele mesmo podem prover valiosa informação para que essa avaliação seja possível.

Uma das rotinas desenvolvidas em Java neste trabalho lê qualquer documento com extensões.txt, .doc e .docx e promove a contagem de cada palavra diferente nesse texto. Outra rotina então faz uso dessa informação e identifica, em um determinado relatório de *clusters* gerado previamente, qual dos termos lidos se encontra em algum *cluster* desse relatório.

Estas rotinas podem ser utilizadas em processos de avaliação psicológica (e em particular em processos seletivos), uma vez que, a partir de um documento que contenha discursos transcritos, redações contendo avaliações de pessoas que conheçam bem os candidatos ao posto de trabalho em questão, transcrições de entrevistas realizadas com esse candidato ou redações elaboradas pelo próprio candidato, as mencionadas rotinas podem ser utilizadas para a produção de um relatório de identificação da presença ou não de traços de interesse nesse documento e suas frequências nesse documento.

A seguir se apresentará um exemplo do uso dessas rotinas na leitura de um documento contendo uma transcrição de uma entrevista (fictícia) com um candidato.

A tabela a seguir apresenta o texto contendo a entrevista.

Tabela 39: Documento ilustrativo de texto contendo entrevista (transcrita) com participante de processo seletivo.

Tell me about yourself.

My background to date has been centered around preparing myself to become the very best skilled professional I can become. Let me tell you specifically how I've worked to develop my personal and technical skills. I spent my whole life trying to better understand myself in face to every conceivable situation. In this process I begun to note the points where I had to change, I had to get better (speciaaly) in those situations whre I tended to lose my humor, or get angry in face of some specific situations. As a corollary I also became more capable of better undertand the others in face of the same situations and thus, more tolerant with them. On the other hand, I spent a lot of time forcing myself to look for the details in every situation and became a very analytical person. Those abilities helped me a lot in the financial area.

So I can say that today I'm very constructive in the senses that I'm looking forward to build and keep new contacts and relationships, I'm always trying to fix things, find new solutions, and always try to see things in the more positive way possible.

Why should I hire you?

Because I sincerely believe that I'm the best person for the job. I realize that there are many other college students who have the ability to do this job. I also have that ability. But I also bring an additional quality that makes me the very best person for the job--my attitude for excellence. Not just giving lip service to excellence, but putting every part of myself into achieving it.

What is your long-range objective? Where do you want to be 10 or 15 years from now?

Although it's certainly difficult to predict things far into the future, I know what direction I want to develop toward. Within five years, I would like to become the very best visionary manager your company has. In fact, my personal career mission statement is to become a world-class leader in the entire industry. I will work toward becoming the expert that others rely upon. And in doing so, I feel I will be fully prepared to take on any greater responsibilities that might be presented in the long term.

How has your education prepared you for your career?

As you will note on my resume, I've taken not only the required core classes in the mathematical field, I've also gone above and beyond. I've taken every class the college has to offer in the field and also completed an independent study project specifically in this area. But it's not just taking the classes to gain academic knowledge--I've taken each class, both inside and outside of my major, with this profession in mind. In addition, I've always tried to keep a practical view of how the information would apply to my job.

Are you a team player?

Very much so. In fact, I've had opportunities in both athletics and academics to develop my skills as a team player. I was involved in basketball and been a strong leading athlete and at the intramural level, including leading my team in assists during the past year--I always try to help others achieve their best. In academics, I've worked on several team projects, serving as both a member and team leader. I've seen the value of working together as a team to achieve a greater goal than any one of us could have achieved individually.

Have you ever had a conflict with a boss or professor? How was it resolved?

Yes, I have had conflicts in the past. Never major ones, but certainly there have been

situations where there was a disagreement that needed to be resolved. I've found that when conflict occurs, it's because of a failure to see both sides of the situation. Therefore, I ask the other person to give me their perspective and at the same time ask that they allow me to fully explain my perspective. At that point, I would work with the person to find out if a compromise could be reached. If not, I would submit to their decision because they are my superior. In the end, you have to be willing to submit yourself to the directives of your superior, whether you're in full agreement or not. An example of this was when I had to deal with work colleague that constantly accused me of being headstrong and mulish.

What is your greatest weakness?

I would say my greatest weakness has been my lack of proper planning in the past. I would overcommit myself with too many variant tasks, then not be able to fully accomplish each as I would like. However, since I've come to recognize that weakness, I've taken steps to correct it. For example, I now carry a planning calendar in my pocket so that I can plan all of my appointments and "to do" items. So I can say I'm a prepared person.

If I were to ask your professors to describe you, what would they say?

I believe they would say I'm a very energetic person, that I put my mind to the task at hand and see to it that it's accomplished. They would say that if they ever had something that needed to be done, I was the person who they could always depend on to see that it was accomplished. They would say that I always took a keen interest in the subjects I was studying and always sought ways to apply the knowledge in real world settings. Am I just guessing that they would say these things? No, in fact, I'm quite certain they would say those things because I have with me several letters of recommendation from my professors, and those are their very words.

What qualities do you feel a successful manager should have?

The key quality should be leadership--the ability to be the visionary for the people who are working under them. The person who can set the course and direction for subordinates. A manager should also be a positive role model for others to follow. The highest calling of a true leader is inspiring others to reach the highest of their abilities. I'd like to tell you about a person who I consider to be a true leader. I refer to Abraham Lincoln, who assumed the presidency on a period of civil war and conducted the contry to good directions.

If you had to live your life over again, what would you change?

That's a good question. I realize that it can be very easy to continually look back and wish that things had been different in the past. But I also realize that things in the past cannot be

changed, that only things in the future can be changed. That's why I continually strive to improve myself each and every day and that's why I'm working to continually increase my knowledge in the management field. That's also the reason why I want to become the very best, prominent and sociable professional your company has ever had. To make positive change. And all of that is still in the future. So in answer to your question, there isn't anything in my past that I would change. I look only to the future to make changes in my life.

Este texto foi alterado em relação a versão disponível no endereço eletrônico acima. Em particular, o texto originalmente disponível recebeu a inclusão de respostas fictícias às perguntas nele presentes. Estas respostas foram dadas como modelos de respostas possíveis e de modo que se pudesse ilustrar utilizações das rotinas desenvolvidas em linguagem Java. Fonte do texto original: Fonte: <http://www.collegegrad.com/interview/jobinterviewstrengthandweakness.shtml>.

A rotina Java, ao ler o documento de texto, e realizar as contagens das frequências de cada traço de interesse (a partir do universo dos 222 inicialmente considerados), produzirá o seguinte relatório, apresentado na tabela abaixo:

Tabela 40: Relatório de identificação da frequência de traços em textos de avaliação de um determinado indivíduo.

<i>Cluster</i>	<i>Traço</i>	<i>Ocorrências</i>
<i>Cluster 1</i>	<i>analytical</i>	1
<i>Cluster 1</i>	<i>prepared</i>	2
<i>Cluster 1</i>	<i>skilled</i>	1
<i>Cluster 3</i>	<i>sociable</i>	1
<i>Cluster 7</i>	<i>visionary</i>	2
<i>Cluster 10</i>	<i>practical</i>	1
<i>Cluster 10</i>	<i>constructive</i>	1
<i>Cluster 11</i>	<i>energetic</i>	1
<i>Cluster 11</i>	<i>strong</i>	1
<i>Cluster 13</i>	<i>academic</i>	1
<i>Cluster 22</i>	<i>leading</i>	2
<i>Cluster 22</i>	<i>prominent</i>	1
<i>Cluster 41</i>	<i>headstrong</i>	1

O relatório apresentado acima, conta as ocorrências de traços pertencentes ao conjunto dos 222 traços (listados no capítulo 2) presentes no texto “*Job Interview Strength and Weakness*”, disponível no Anexo deste trabalho; mais particularmente nas respostas dadas pelo candidato fictício às perguntas feitas em uma entrevista (e depois transcritas em um documento de texto para que as rotinas em questão pudessem ser utilizadas) presentes no mesmo.

Este texto apresenta respostas ilustrativas a perguntas prováveis de serem feitas em um processo seletivo interno ou externo. As contagens de ocorrências dos 222 traços neste texto se deram somente a partir das seções contendo as respostas do candidato fictício, ignorando-se as demais seções.

Com base na presença ou ausência de traços de interesse em documentos desse tipo, obtêm dados adicionais que corroboram na avaliação inicial do indivíduo sendo analisado, o que será considerado para a continuidade ou não do mesmo em um determinado processo seletivo.

8.3.1.1.2 Procedimentos de avaliação sistemática e melhorias individuais dos funcionários.

Os processos de avaliação psicológica nas diversas organizações humanas também podem visar às melhorias gradativas dos serviços providos por estes membros.

Assim, por exemplo, na área de segurança pública, um departamento policial pode estar empenhado em realizar avaliações psicológicas frequentes de seus policiais visando garantir a qualidade de seus serviços, sua aptidão para suas funções e a segurança da população.

Um gerente de departamento pode estar interessado na melhoria contínua de seus funcionários (do ponto de vista de suas características pessoais) e de que os mesmos se engajem em esforços de melhorias contínuas em relação a determinados aspectos de suas personalidades.

Nestes contextos, processos de avaliações psicológicas podem se dar por diversos meios: *feedbacks* de colegas de trabalho, entrevistas, redações etc. Desde que estes relatos estejam na forma de documento de texto, as ferramentas desenvolvidas aqui, como visto, podem auxiliar no esforço de avaliação psicológica dos indivíduos de qualquer organização.

8.3.1.1.3 Formação de grupos de trabalho afins.

Finalmente, os métodos de avaliação psicológica que podem ser implementados com o auxílio das ferramentas desenvolvidas neste trabalho podem visar a formação de grupos de trabalho afins.

Assim, um grupo de trabalho pode ser concebido em termos do número de seus participantes, das capacidades técnicas de cada um e da personalidade ideal necessária para cada posição neste grupo.

Deste modo, as ferramentas desenvolvidas neste trabalho, ao fornecerem instrumentos que visem também a facilitar a tarefa de avaliação psicológica de cada indivíduo, também ajudam indiretamente na tarefa de formação de grupos de trabalho afins.

8.3.1.2 Aplicações clínicas

Além da possibilidade de avaliações psicológicas por meio de *ratings* numéricos, atribuídos a traços de personalidade organizados em *clusters*, desenvolveu-se outro conjunto de funcionalidades que podem ser utilizadas em aplicações clínicas.

Este conjunto de funcionalidades visa:

- auxiliar o clínico na avaliação psicológica do paciente,
- auxiliar o clínico na avaliação do progresso do paciente,
- auxiliar o clínico na estruturação de relatórios sobre um determinado paciente.

Uma vez que a prática clínica envolve as atividades de elaboração de relatos e relatórios sobre pacientes e a leitura e interpretação de relatos e redações elaborados por pacientes, se desenvolveu algumas aplicações que visam ajudar o profissional envolvido nestas tarefas.

Sabe-se que, uma das formas de se proceder a avaliação psicológica de indivíduos é, além da análise de resultados de questionários, a análise de relatos desses indivíduos. Através dessa análise, pode-se perceber as ocorrências de palavras e de termos com as quais é possível se avaliar este indivíduo.

Deste modo, as frequências de ocorrências de termos e de traços também podem auxiliar o clínico no seu processo de avaliação e/ou de terapia.

Assim, um método que promova a contagem automática das palavras presentes nos relatos e/ou nas redações de um indivíduo pode auxiliar a prática clínica. Mais ainda, quando os termos de interesse para o clínico (com os quais deseja observar suas frequências de ocorrência nestes) relatos e/ou redações estão organizados em *clusters* úteis para este profissional, estas ferramentas tornam-se ainda mais atrativas; pois:

- auxiliam o profissional no processo de avaliação dos pacientes e seus progressos;
- organizam informações úteis para o profissional e
- poupam tempo do profissional na elaboração destes relatórios.

Assim, funcionalidades que auxiliem profissionais desta área a atingir estes objetivos tendem a ser úteis no ramo.

Assim, algumas das rotinas desenvolvidas em Java promovem a importação de textos em qualquer formato, nas extensões.txt, doc, docx e.xml e a contagem de cada palavra presente no texto. A seguir as rotinas desta interface identificam quais das palavras presentes no texto estão presentes no relatório de *clusters* de palavras no *thesaurus*, previamente gerado e importado por esta interface.

Como visto, após esta identificação, as contagens efetuadas para estas palavras são postas do lado de cada palavra do relatório de *clusters*.

Como estas rotinas auxiliam o trabalho clínico?

Estes expedientes auxiliam o trabalho clínico, pois:

- Redações e demais textos produzidos por pacientes podem ser transformados em documentos de texto nos formatos acima mencionados. Neste sentido, se a redação for feita com letra de forma, *softwares* OCR podem automaticamente transformá-los em textos formatados. A partir desse passo podem-se utilizar as funcionalidades construídas no auxílio da prática clínica.

- Anotações feitas pelo clínico a cerca das dinâmicas com o paciente ou ainda seus relatórios sobre o mesmo podem ser transformadas em texto formatado pelo mesmo procedimento mencionado (utilizando-se *softwares* OCR ou ainda por digitação do documento físico em um editor de texto).

- Finalmente, a gravação do áudio das sessões com pacientes também podem ser transformadas em texto; o que tornaria novamente os desenvolvimentos do sistema úteis para a prática clínica.

Adicionalmente, as ferramentas desenvolvidas também podem auxiliar o profissional da área clínica na elaboração de seus relatórios, fornecendo ao mesmo um conjunto de termos relacionados, com os quais o profissional pode elaborar de forma mais precisa e fidedigna as avaliações dos pacientes (pela escolha dos termos que melhor descrevam e traduzam suas percepções acerca do paciente, de suas características atuais, seus históricos e seus progressos).

Por fim, as ferramentas desenvolvidas também podem auxiliar o profissional da área na elaboração de seus relatórios. Isto se dá por meio dos relatórios de *clusters*, onde os traços de interesse estão agrupados em *clusters* de acordo com a linguagem natural, e onde estes *clusters* são determinados pelo algoritmo de clusterização MCL.

Assim, a visualização deste relatório pode sugerir termos mais apropriados para serem utilizados na descrição mais fidedigna e precisa possível do paciente (e de seu progresso) pelo clínico, haja vista que em cada *cluster* tem-se um conjunto de termos relacionados entre si e por meio destes agrupamentos pode-se escolher os termos mais apropriados para essas descrições.

Da mesma forma, a possibilidade de visualização do grafo relacionado a um conjunto de termos de interesse também auxilia nesta tarefa. Assim, embora a ferramenta de

visualização de grafos utilizada (*software* R com pacote *Igraph*) não seja um desenvolvimento deste trabalho, ela pode auxiliar também o profissional nestas tarefas (embora demandando do mesmo a elaboração de linhas de comando).

8.3.1.3 Aplicações na área de pesquisa em teoria de traços de personalidade

Como anteriormente apresentado alguns dos principais estudos na área de traços de personalidade foram estudos psicoléxicos; estudos que se valiam da hipótese lexicográfica (que, como visto, enuncia que as diferenças individuais mais socialmente relevantes serão codificadas na linguagem natural), que continua a ser amplamente adotada e aceita por parte dos pesquisadores na área.

(Como apresentado na seção relativa às motivações do estudo), mesmo o modelo mais aceito na literatura atualmente para explicar a personalidade humana (o modelo dos Cinco Grandes Fatores da personalidade de GOLDBERG (1993), como visto, apresenta problemas ainda não resolvidos).

Ainda, de acordo com JACKSON e PAUNONEN (2000) a pesquisa na área de personalidade ainda está aberta a contribuições e pesquisas.

Retornando aos estudos psicoléxicos anteriormente feitos, notou-se que diversos deles utilizaram-se da coleta de termos de um dicionário (físico) efetuados pelo estudo de ALLPORT e ODBERT (1936), como exposto no capítulo, seguida de classificação desses traços coletados em categorias onde os mesmos tivessem maior similaridade entre si no que se refere aos aspectos comportamentais aos quais se referiam.

Como visto no capítulo 4 deste trabalho, pelo fato de existirem muitas palavras descritoras da personalidade, muitas delas possuem significados similares; isso implica que muitos dos TP se sobrepõem uns aos outros quanto a esses significados. Tais TP podem ser agrupados de forma a definirem aspectos (ou dimensões) amplos da personalidade; o que de fato é uma etapa fundamental para um modelo que vise explicar, com o menor número de variáveis possíveis os comportamentos humanos.

A esse respeito, nada mais natural que este esforço de agrupamento de traços com significados similares seja auxiliado por um agrupamento dos mesmos segundo as similaridades de sentido que assumem de acordo com a linguagem natural. Ocorre que, os CPT representam exatamente esta possibilidade (como pode ser constatado pelas suas propriedades, já mencionadas anteriormente neste trabalho) e, portanto as ferramentas computacionais implementadas neste trabalho podem auxiliar novas pesquisas que se

valham de uma coleta exaustiva de traços em um dicionário seguida de posterior classificação desses traços segundo essa similaridade.

A esse respeito, como já se conhece na literatura uma enorme quantidade de termos descritores de comportamento, a busca exaustiva desses termos por meio da leitura de um dicionário (ou outra base de dados) pode ser facilitada por meio da ferramenta implementada em linguagem Java, que checa, no *thesaurus* se um determinado conjunto de palavras de interesse existe no mesmo.

Os algoritmos de clusterização implementados nesta tese podem fornecer uma organização dos traços selecionados, em uma determinada aplicação, de acordo com a similaridade entre eles no referente aos comportamentos por eles representados.

Neste sentido, estes mesmos *clusters* podem auxiliar o pesquisador na busca de termos descritores de personalidade que sejam os mais abrangentes possíveis (no que se refere aos aspectos comportamentais a que se referem).

Deste modo, como a área de traços ainda apresenta uma série de desafios, pesquisas que visem superar as limitações atuais (como por exemplo, as mencionadas com o modelo de cinco grandes fatores) é bem vinda. Estas pesquisas podem incluir, por exemplo, a construção de modelos de grandes fatores a partir de dicionários e bases de dados específicas.

Para tanto, ferramentas que possam processar grandes conjuntos de palavras em *thesaurus* digitais podem significar não só uma forma rica de coleta de termos descritores de personalidade; como também os *clusters* de termos nesses bancos de palavras podem fornecer um método inicial de organização e classificação de termos para estes pesquisadores. As ferramentas implementadas em código Java atendem estes pré-requisitos.

Por outro lado as ferramentas desenvolvidas nesta tese tornam-se também importantes e relevantes para estes tipos de aplicação uma vez que se constata que a língua, como uma das principais expressões da cultura de uma população, é dinâmica, variando ao longo do tempo no que diz respeito aos sentidos que determinados vocábulos (e expressões) ganham. Adicionalmente novas palavras surgem e outras se tornam obsoletas. Assim pesquisas psicoléxicas podem ser repetidas após grandes intervalos de tempo para darem conta dessas mudanças na língua natural e seria desejável se ter ferramentas que agilizassem os procedimentos de coleta e organização de termos.

8.3.2 Outras aplicações

8.3.2.1 Aplicações linguísticas e ensino de línguas

Em aplicações linguísticas pode-se estar interessado, por exemplo, na busca de sinônimos de palavras.

As ferramentas implementadas neste trabalho também podem auxiliar o pesquisador linguista nestas tarefas. Como visto anteriormente, boa parte das associações (diretas) entre palavras em um *thesaurus*, revelam palavras cumprindo o papel de sinônimas umas das outras.

Assim, relatórios que revelem *clusters* de palavras podem facilitar o linguista na busca por sinônimos de palavras; uma vez este relatório informa agrupamentos de palavras de acordo com suas proximidades umas das outras, no *thesaurus*.

Do mesmo modo, e como ressaltam HIGGISON e PHLEA (1986) tais *clusters* podem auxiliar no ensino e aprendizagem de línguas, principalmente no desenvolvimento de vocabulário, uma vez que os termos que os termos vizinhos uns dos outros, em um dado *cluster* qualquer, geralmente são sinônimos entre si; e também porque os sentidos gerais dos *clusters*, se relacionando às palavras neles contidos, podem ser empregados no ensino de vocabulário por temas, isto é, temático.

8.3.2.2 Aplicações na área de *Knowledge Management* (especialmente organização de documentos).

Aplicações de *clusters* de palavras em *thesaurus* também podem auxiliar no desenvolvimento de algoritmos que visem classificar documentos com base no grau de similaridade de seus conteúdos.

Assim, com base nas frequências de palavras em diferentes textos podem-se criar critérios para se julgar o grau de similaridade (ou de dissimilaridade) de seus conteúdos. Neste sentido, a ferramenta que promove a contagem de palavras em documentos de texto (em diversas extensões) pode auxiliar nas elaborações dessas medidas de similaridade de conteúdos de texto.

Mais ainda, como os relatórios de contagem de texto produzidos pelo sistema, associam também cada palavra a um *cluster* e cada *cluster* possui um sentido amplo ligando todas as suas palavras constituintes, pode-se utilizar esses sentidos gerais dos *clusters* como macro-temas ou macro assuntos que os textos podem possuir ou não.

Deste modo, estes relatórios podem auxiliar nos procedimentos de organização de documentos e, portanto, na gestão do conhecimento.

8.3.2.3 Outras aplicações em Teoria de Grafos/Algoritmos em Grafos

Como levantado anteriormente, todas as funcionalidades relativas a algoritmos em grafos e de clusterização em grafos estão disponíveis para serem aplicadas em problemas de grafos em geral (desde que se trate de 1-grafos). Esta possibilidade decorre do fato de que todas as rotinas implementadas em linguagem Java dependem, direta ou indiretamente, somente de um tipo de relatório: a lista de adjacência do grafo.

Como ressaltado, esta lista pode ser produzida nas rotinas implementadas nessa linguagem (ou no sistema em PHP) ou ser digitada pelo usuário; desde que respeite os formatos requeridos na importação por estas aplicações. Isto permite que os algoritmos implementados sejam utilizados em diversos problemas envolvendo algoritmos em grafos.

CAPÍTULO 9: CONCLUSÕES E RECOMENDAÇÕES DE ESTUDOS FUTUROS

9.1 CONCLUSÕES

Este trabalho, ao abordar as propriedades e a importância do problema de determinação de *clusters* de palavras provindas de *thesauri* focou em aplicabilidades dessas estruturas nas áreas clínicas, de pesquisa em traços de personalidade e nas aplicações organizações nas quais podem ser utilizadas (sobretudo em processos seletivos).

Os objetivos deste trabalho foram os seguintes:

- Contribuir para a facilitação da pesquisa em traços de personalidade, práticas clínicas e aplicações organizacionais que usassem traços de personalidade, por meio de soluções algorítmicas integradas que fornecessem flexibilidade de trabalho aos profissionais das áreas supracitadas.

- Desenvolver rotinas que pudessem ser utilizadas na produção de relatórios relacionadas a grafos (não direcionados) de uma forma em geral e no fornecimento de um algoritmo que permitisse a codificação de vértices desse grafo de forma que pudesse servir de *input* à rede neural SOM para a clusterização do mesmo, em problemas onde só se dispusesse de informações relativas às adjacências dos vértices do grafo.

Com base nestes objetivos, implementou-se uma série de rotinas computacionais (e desenvolveu um *software*) capazes de produzir relatórios relacionados a esses agrupamentos com aplicabilidades nas supracitadas áreas.

Essas rotinas, como se pôde constatar, foram capazes de gerar estes relatórios a partir da importação de palavras de um *True Term Thesaurus Database* (TTT).

A seguir se averiguará se os mesmos foram atingidos.

Com relação ao primeiro objetivo, pode-se verificar que o mesmo foi atingido. As rotinas implementadas em Java, PHP, MATLAB e R foram capazes de gerar relatórios relacionados ao grafo subjacente às relações entre palavras (e em particular traços) selecionadas do *thesaurus* utilizado. Abordando-se tal rede de relacionamentos como um grafo, foi possível se construir as seguintes rotinas:

- Rotinas geradoras de listas de adjacência (em PHP e Java) do grafo, a partir do TTT utilizado.

- Rotinas geradoras de matrizes de adjacência (em PHP e Java) do grafo, a partir do citado *thesaurus*.

- Rotinas geradoras de vetores codificando cada palavra utilizada (em PHP) do grafo, a partir do citado banco.

- Rotinas geradoras do relatório de relação entre cada palavra e seu respectivo rótulo numérico (em PHP).

- Rotinas de clusterização em grafos, pelo algoritmo MCL (em Java).

- Rotinas de leitura de arquivos de texto em geral, nas extensões.txt.doc e .docx. Estas rotinas realizam a checagem da presença ou não de cada termo (no caso, traço de personalidade) de interesse em documentos com essas extensões e registram a frequência de cada um deles nesses documentos; identificando cada termo neles presentes nos *clusters* de palavras já definidos anteriormente.

- Rotinas que permitem a checagem se determinadas palavras de interesse se encontram no TTT e seleção de somente estes termos na confecção dos demais relatórios (em linguagem Java).

Como visto anteriormente, estas rotinas proporcionam flexibilidade de trabalho para pesquisadores e profissionais (incluindo-se aqui a possibilidade de geração de CTP e suas visualizações), pois permitem que os mesmos escolham que traços serão incluídos nos relatórios, identifiquem *clusters* a partir dessas escolhas e gerem os demais relatórios baseados nesses agrupamentos.

Em particular para profissionais envolvidos com processos seletivos e avaliação psicológica, as rotinas desenvolvidas, além de permitirem a escolha de traços de interesse para um dado processo, permite a geração de questionários a partir dos mesmos e o cômputo de *ratings* finais para cada profissional/candidato em relação a cada traço, da possibilidade de atribuição de pesos diferenciados a cada um desses traços e a cada *cluster* identificado.

Os profissionais das áreas clínicas também podem utilizar as rotinas aqui implementadas com os propósitos de avaliação de seus pacientes.

Os pesquisadores da área de traços de personalidade e de modelos de personalidade também podem se utilizar das rotinas implementadas, pois as mesmas permitem aos mesmos selecionarem e gerarem os diversos relatórios mencionados a partir de grande quantidade de traços (podendo até mesmo envolver todos os traços) presentes no TTT; assim como a possibilidade de obtenção de uma organização desses traços segundo a linguagem natural.

Como visto, os CTP proporcionam uma forma útil de organização desses traços, uma vez que nestes agrupamentos, termos diretamente relacionados uns aos outros cumprem, na grande maioria dos casos, o papel de sinônimos e, portanto traços diretamente relacionados

entre si tenderão, na maior parte dos casos, a representarem comportamentos muito similares ou idênticos.

Adicionalmente, nestes *clusters*, é sempre possível se identificar sentidos gerais, unindo os termos contidos em cada um desses agrupamentos, o que pode facilitar o esquema de classificação desses traços para posteriores etapas de desenvolvimento de modelos de personalidade.

Como também apresentado, as rotinas desenvolvidas em Java também se prestam a aplicações de grafos não necessariamente associadas ao *thesaurus* aqui adotado. Deste modo, diversas rotinas e relatórios relacionados a grafos não direcionados podem ser utilizadas em aplicações genéricas da Teoria de Grafos.

Todas estas funcionalidades, além de oferecer soluções flexíveis a cada uma das categorias profissionais acima mencionadas, corroboram no sentido de poupar tempo dos mesmos na execução de tarefas mais operacionais (como coleta exaustiva de traços de personalidade e a sua catalogação em categorias, feita imediatamente após esta coleta).

Adicionalmente, algumas destas rotinas estão disponíveis em um sistema de interfaces gráficas (GUI's do sistema PHP) de fácil uso para usuários sem habilidades de desenvolvimento e programação, o que se constitui um atrativo adicional para tais usuários.

Objetivo 2: Desenvolver um algoritmo de codificação dos vértices de um grafo em vetores, que possam permitir o uso de redes neurais SOM em problemas de clusterização de grafos.

Como visto, o problema de clusterização abordado possui peculiaridades que motivaram o desenvolvimento do algoritmo de codificação dos vértices do grafo subjacente à rede de relacionamentos entre palavras no TTT. Neste problema, não se dispunha de nenhuma informação inicial a cerca do mencionado grafo que não as adjacências entre seus vértices.

Para que esta rede neural pudesse ser utilizada, fazia-se necessário codificar tais vértices como vetores, de modo que suas coordenadas informassem algo sobre as distâncias entre os primeiros (o que se verificou possível quando a distância euclidiana e o seu quadrado fossem computados a partir das coordenadas desses vetores).

Como pôde ser constatado, quando se computa o quadrado das distâncias euclidianas entre vetores resultantes da codificação implementada pelo algoritmo apresentado, efetuada em grafos do tipo árvore, se obtém as exatas menores distâncias entre os vértices por esses vetores.

Resultado idêntico se obtém aplicando-se o algoritmo em florestas; com a única diferença de que distâncias explosivas serão atribuídas entre pares de vértices pertencentes a diferentes componentes conexas do grafo.

Quando se aplica este algoritmo em grafos cíclicos, se obtém aproximações das menores distâncias entre pares de vértices pertencentes a ciclos ao passo que o quadrado das distâncias euclidianas computadas entre vetores representando pares de vértices pertencentes a regiões não cíclicas do grafo continuam reproduzindo as exatas menores distâncias entre esses vértices.

O algoritmo de codificação aqui apresentado se presta a aplicações de clusterização de grafos cíclicos, pois é capaz de atribuir distâncias que reflitam as estruturas de ciclos do grafo (atribuindo distâncias maiores entre vértices pertencentes a diferentes regiões cíclicas do grafo do que entre vértices pertencentes a uma mesma região cíclica).

Este algoritmo apresentou excelentes resultados, quando empregado em conjunto com redes neurais SOM na clusterização de vértices de diferentes grafos testados, incluindo o grafo associado ao conjunto de 222 traços selecionados para os desenvolvimentos computacionais desta tese.

Este algoritmo tem o mérito de ser um algoritmo simples de ser implementado e compreendido (e principalmente de poder tratar o problema de clusterização abordado com redes SOM). Este algoritmo se constitui em uma contribuição original na área de redes neurais, algoritmos discretos e clusterização.

Na área de *softwares*, também proporcionou outro desenvolvimento inédito: o desenvolvimento de um programa com interface simples ao usuário, capaz de ler o TTT e gerar alguns relatórios de grafos (subjacentes aos relacionamentos entre conjuntos de palavras de interesse). Mais especificamente, os relatórios de matrizes de adjacência, listas de adjacência e relatório de relação entre palavras e rótulos numéricos; desenvolvido em linguagem PHP; o que é extremamente importante para desenvolvimentos baseados em algoritmos de grafos em tal *thesaurus*. Também executa o referido algoritmo de codificação de vértices de um grafo em vetores.

Este trabalho também fez outras contribuições originais: aplicou de forma inédita redes SOM (conjuntamente com o acima mencionado algoritmo de codificação de vértices) em um problema de clusterização envolvendo palavras em um *thesaurus*. Também não se encontrou nenhum trabalho utilizando o *thesaurus* aqui adotado.

Os resultados de clusterização obtidos com esta rede neural e com a citada metodologia de codificação foram muito satisfatórios. Esta metodologia foi capaz de encontrar agrupamentos de traços bem consistentes (no sentido de que cada um desses agrupamentos contém traços intimamente relacionados entre si, e agrupados segundo sentidos mais amplos, comuns a todos os traços alocados em um mesmo *cluster*).

Resultados similares se verificaram com os *clusters* determinados pela metodologia *Markov Clustering* (MCL).

A respeito dos resultados obtidos com essas duas metodologias de clusterização, no conjunto de 222 traços, verificou-se que o algoritmo MCL encontrou menos agrupamentos do que a rede SOM. De uma forma geral estes agrupamentos contiveram mais palavras do que os encontrados por esta última metodologia.

Ao se comparar os resultados de clusterização obtidos com ambas as metodologias com as variáveis e fatores do modelo 12PF de Raymond Cattell, se observou certa correspondência entre algumas variáveis/fatores desse modelo e os *clusters* identificados pelos algoritmos MCL e redes SOM.

Estas correspondências se traduziram no fato de em várias variáveis (e conseqüentemente nos fatores das quais as mesmas fazem parte) do modelo se observou que alguns de seus traços componentes se encontravam alocados no mesmo *cluster* identificado pelos algoritmos MCL (resultados similares se verificando em relação aos *clusters* identificados pela rede SOM) ou em *clusters* localizados em regiões próximas do grafo associado à rede de ligações entre os 222 traços utilizados.

Como as variáveis desse modelo foram obtidas por meio das correlações verificadas entre os *ratings* obtidos nas respostas de questionários envolvendo estes traços e os *clusters* obtidos pelos algoritmos MCL/SOM foram obtidos de acordo com a similaridade entre estes termos segundo a linguagem comum, verificou-se que ambos os métodos de obtenção de *clusters* produziram, para algumas variáveis do modelo, resultados similares (no sentido de abrigarem alguns traços em comum).

Deve-se agora retomar as questões que esta tese se propunha a responder. As mesmas são:

- Podem algoritmos em grafos, em particular de clusterização, auxiliarem na pesquisa e prática relacionadas à personalidade humana?

- É possível oferecer métodos que possibilitem a utilização de algoritmos de clusterização conhecidos como redes neurais de Kohonen para serem aplicados nos problemas de clusterização em grafos?

- Podem as técnicas aqui desenvolvidas formar uma base para o desenvolvimento de uma ferramenta gerencial que auxilie nas diversas funções organizacionais envolvendo *personality assessment*?

Pelas considerações feitas neste capítulo, pode-se de imediato reconhecer que estas três perguntas foram respondidas.

9.2 RECOMENDAÇÕES DE ESTUDOS FUTUROS

Seria prudente reservar para estudos futuros algum ponto que tenha sido abordado ao longo deste trabalho.

Um ponto relevante que deve ser dedicado para estudos futuros se refere à modificação do TTT, com a ajuda de um linguista, para abrigar os exatos 171 traços propostos por Cattell e a seguir, se efetuar novamente os experimentos de clusterização com os dois algoritmos aqui empregados.

Uma vez obtidos os *clusters* finais obtidos com estes 171 traços e com o conjunto menor de 105 utilizados por Cattell para construção de seu modelo 12PF, seria extremamente interessante e relevante se proceder à comparação destes *clusters* com as variáveis e fatores desse modelo. Pelo fato de alguns desses 171 traços não estarem disponíveis no TTT utilizado e de se ter optado pela inclusão de palavras candidatas a substitutas a esses termos ausentes do thesaurus (ao invés de se alterar o thesaurus para abrigá-las, o que demandaria a participação de linguistas), este trabalho acabou implementando experimentos de clusterização com um conjunto de traços um pouco diferente do conjunto utilizado por Cattell, de modo que uma comparação plenamente informativa deveria ser implementada como mencionado.

Como extensão das comparações mencionadas, seria também interessante se comparar os *clusters* obtidos com as duas metodologias de clusterização a partir dos mencionados 105 traços com os demais modelos de Cattell (o modelo de 16 fatores e o modelo com as cinco grandes dimensões da personalidade) assim como os demais modelos de cinco fatores da personalidade de outros autores. Estas comparações buscariam evidências a favor (ou contra) a possibilidade de tais *clusters* obtidos no thesaurus, conseguirem (ou não) reproduzir os fatores desses modelos.

Outros temas objetos de estudos futuros incluem o desenvolvimento de interfaces gráficas (GUI's) de fácil utilização das rotinas implementadas em linguagem Java.

Sugere-se também, se utilizar os algoritmos de clusterização aqui empregados em outras bases de dados e em experimentos concretos de desenvolvimento de modelos de personalidade.

REFERÊNCIAS BIBLIOGRÁFICAS

ALLPORT, G. W., ODBERT, H. S., 1936, "Trait-names, a psycho-lexical study" *Psychological Monographs.*, v.47, Pp. 1-171, Harvard Psychological Laboratory, Harvard, Disponível em: <

http://psych.colorado.edu/~carey/Courses/PSYC5112/Readings/psnTraitNames_Allport.pdf >. Acesso em: 01 Dez 2011.

ALTHOFF, R. W., 2010, "The Big Five Personality Traits as Predictors of Academic Maturity". Dissertação de M. SC., Graduate School of Eastern Illinois University, Illinois, Chicago, Estados Unidos. Disponível em: <
<http://thekeep.eiu.edu/cgi/viewcontent.cgi?article=1604&context=theses> >. Acesso em: 01 set. 2012.

COLLAZO, R. A. , 2009, Aplicação de "Support Vector Machines" à Classificação do Risco de Morte de Pacientes com Síndrome Coronariana Aguda. Dissertação de de M. SC., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

ANKERST, M., BREUNIG, M. M., KRIEGEL, H. P., SANDER, J., 1999, "OPTICS: Ordering Points To Identify the Clustering Structure", *ACM SIGMOID International Conference on Management of Data*, páginas 49-60, Melbourne, AU, ACM Press, Nova York. Disponível em: <: <http://www.cs.uiuc.edu/class/fa05/cs591han/papers/ankerst99.pdf> >. Acesso em: 01 mai. 2013.

ANÔNIMO, 2013, *Material Sobre Clusterização em Grafos Direcionados Acíclicos*, Apostila da Penn State University, Pensylvania. Disponível em: <
<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=A55EA75150C543E74AC1E6140A641424?doi=10.1.1.48.71&rep=rep1&type=pdf> >. Acesso em: 01 nov. 2013.

ANÔNIMO, 2014, Psychometrics for Selection & Development. Disponível em: <
<http://www.linkedin.com/groups/Are-psychometric-tests-useful-in-2249553.S.95668917> >. Acesso em: 23 fev. 2014.

APACHE SERVER. Disponível em: < <http://www.apache.org/> >. Acesso em: 12. abril.2011.

APACHE SERVER COMPILAÇÃO 64 BITS. Disponível em: <
<http://www.apachelounge.com/download/win64/> >. Acesso em: 23. abril.2012.

ARSHAVA, I., NOSENKO, D., NOSENKO, G., SAYEVICH, A., 2013, "The Role of Simultaneous Pursuing of Academic and Athletic Activities over Youth in Personality

Development”, *Psychology*, v. 4 n.3A, (Mar), pp. 309-317. Disponível em: <: <http://www.scirp.org/journal/PaperDownload.aspx?paperID=29254> >. Acesso em: 01 mai. 2013.

ARRUDA, E., F., 2010, *Processos de Decisão Markovianos: Algoritmos e Aceleração de Convergência*, Apostila da Faculdade de Engenharia da PUCRS – Pontifícia Universidade Católica do RS. Disponível em: < http://www.pucrs.br/edipucrs/XISalaoIC/Engenharias/Engenharia_Eletrica/82963-LEANDROACACIOESVAEL.pdf >. Acesso em: 15 mai. 2014.

BAKER, L. D., MC CALLUM, A. K., , 1998, “Distributional Clustering of Words for Text Classification”, *Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval*, páginas 96-103, Melbourne, AU, ACM Press. Disponível em: <: <http://blondie.cs.byu.edu/CS652/baker98distributional.pdf> >. Acesso em: 01 mai. 2012.

BALAHUR, A., HERMIDA, J., , 2012, “Extending the EmotiNet Knowledge Base to Improve the Automatic Detection of Implicitly Expressed Emotions from Text”, *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pp. 1207-1214. Disponível em: <: http://www.lrec-conf.org/proceedings/lrec2012/pdf/945_Paper.pdf >. Acesso em: 01 mai. 2012.

BEALE, M., H., HAGAN, M. T., DEMUTH, H. B., *Neural Network Toolbox User's Guide R2014a*, 2014, 1ed, New England, The MathWorks, Inc.. Disponível em: < <http://www.mathworks.com/> >. Acesso em: 14 de janeiro. 2014.

BENOIT, L., J., 2012, *Psychological Report*, Lafayette, LA. Disponível em: < <http://www.katc.com/files/LAVERGNE%20REPORT.pdf> >. Acesso em: 01 mar 2013.

BERTSEKAS, DIMITRI P., 2003, *Nonlinear Programming*. 2 ed. Belmont, Athena Scientific.

BISHOP, C. M., 1995, *Neural Networks for Pattern Recognition*. 1 ed. Oxford, Clarendon Press.

- BOUCHARD JR., T., J., 2004, “Genetic Influence on Human Psychological Traits”, American Psychological Society, v.13, n.4, pp. 148-151. Disponível em: < <http://stormchan.org/study/src/1347441918244.pdf> >. Acesso em: 01 set. 2012.
- BRASLAVSKI, P., ALSHANSKI, G., SHISHKIN, A., “ProThes: Thesaurus-Based Meta-Search Engine for a Specific Application Domain”, 2004, Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, pp. 222-223, USA, ACM Press, Nova York. Disponível em: <: <http://www2004.wwwconference.org/docs/2p222.pdf> >. Acesso em: 03 mai. 2013.
- BAKER, L. D., MC CALLUM, A. K., , 1998, “Distributional Clustering of Words for Text Classification”, *Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval*, páginas 96-103, Melbourne, AU, ACM Press, Nova York. Disponível em: <: <http://blondie.cs.byu.edu/CS652/baker98distributional.pdf> >. Acesso em: 01 mai. 2012.
- CALÔBA, L., P., 2009, *Uma Pequena Introdução às Redes Neurais Artificiais*, Apostila do IX CBRN/IC, Ouro Preto, Minas Gerais. Disponível em: < http://www.eletrica.ufpr.br/anais/cbrn/2009/minicurso_Caloba.pdf >. Acesso em: 01 set. 2011.
- CALÔBA, L., P., RAJAGOPAL, R., SEIXAS, J., M., 1999, “Redes Neurais Aplicadas na Separação de partículas em Calorímetros Cintilantes”, *Proceedings of the IV Brazilian Conference on Neural Networks*, (Jul), pp.257-262, , São José dos campos, São Paulo. Disponível em: < http://www.ele.ita.br/cnrrn/artigos-4cbrn/4cbrn_060.pdf >. Acesso em: 15 set. 2011.
- CARVALHO JR., J. G., 2002, *Modelos Conexionistas – Redes Neurais*, Apostila da PUC-Rio, Rio de Janeiro, PUC.
- CASTRO, E., VLADIMIR., 2002, “Why So Many Clustering Algorithms – A Position Paper”, *ACM SIGKDD Explorations Newsletter*, v. 4 n.1, (Jun), pp. 65-75. Disponível em: <: <http://www.kdd.org/sites/default/files/issues/4-1-2002-06/estivill.pdf> >. Acesso em: 01 mai. 2012.
- CATTELL, H., E. P., MEAD, A. D., 2007, “16 Personality Factor Questionnaire (16PF)”, In: BOYLE, G.,J., MATTHEWS, G., SAKLOFSKE, D.H. (Eds.), *Handbook of*

Personality Theory and Testing: Vol 2: Personality Measurement and Assessment. Ed. Sage, London. Disponível em: <
<http://people.wku.edu/richard.miller/520%2016PF%20Cattell%20and%20Mead.pdf> >. Acesso em: 01 set. 2012.

- CATELL, R. B., 1943a, "The Description of Personality I: Foundations of Trait Measurement", *Psychological Review*, v. 50 n.6, (Nov), pp. 559-594.
- _____, 1943b, "The Description of Personality II: Basic Traits Resolved Into Clusters", *Journal of Abnormal and Social Psychology*, v. 38 n.4, (Out), pp. 476-506.
- _____, 1944a, "A Note on Correlation Clusters and Cluster Search Methods", *Psychometrika*, v.9, n.3, (Set), pp. 169-184.
- _____, 1944b, "Parallel Proportional Profiles and Other Principles for Determining the Choice of Factors by Rotation", *Psychometrika*, v.9, n.4, (Dez), pp. 267-288.
- _____, 1945a, "The Principal Trait Clusters for Describing Personality", *Psychological Bulletin*, v.42, n.3, (Mar), pp. 129-161.
- _____, 1945b, "The Description of Personality: Principles and Findings in a Factor Analysis", *The American Journal of Psychology*, v.58, n.1, (Jan), pp. 69-90.
- _____, 1946a, *Description and Measurement of Personality*, Journal of Abnormal and Social Psychology, 1 ed., Londres, George G. Haarp, W. C. 1.
- _____, 1946b, "Personality Structure and Measurement I: the Operational Determination of Trait Unities", *The British Journal of Psychology*, v.1, n.36, (Dez), pp. 88-103.
- _____, 1946c, "Personality Structure and Measurement II: the Determination and Utility of Trait Modality", *The British Journal of Psychology*, v.1, n.36, (Dez), pp. 159-174.
- _____, 1947, "Confirmation and Clarification of Primary Personality Factors", *Psychometrika*, v.12, n.3, (Set), pp. 197-220.
- _____, 1956, "A Shortened "Basic English" Version (From C) of the 16 PF Questionnaire", *The Journal of Social Psychology*, v.1, n.44, pp. 257-278. Disponível em: <
http://psych.colorado.edu/~carey/courses/psyc5112/readings/psnstructure_cattell.pdf >. Acesso em: 01 mai. 2012.
- _____, 1957, "Personality and Motivation: Structure and measurement", *The British Journal of Psychology*, v.1, n.36, (Dez), pp. 159-174. In: Big Five Personality Traits, 2012.

- _____, 1970, Handbook for the Sixteen Personality Factor Questionnaire, Institute for Personality and Ability Testing, Champaign, IL. In: MATTHEWS, G. et al, 2003, p.18.
- _____, 2008, *The Scientific Analysis of Personality*, 1 ed., Nova Brunswick, Aldine Transaction.
- CATTELL, R., B., CATTELL, K, CATTELL H., 2000, “16 Personality Factor Questionnaire (B)”, AustPsychPress. Disponível em: < http://austpsychpress.com.au/pdf_reports/Catalogues/16PF.pdf >. Acesso em: 01 set. 2012.
- CAVICCHIOLI, G., 2008, “Estilos de Liderança de Fundadores de Empresas Familiares do Setor de Transporte Rodoviário de Cargas no Brasil e a Continuidade do Negócio”, M. SC., PUC – SP, Rio de Janeiro, RJ, Rio de Janeiro, RJ, Brasil. Disponível em: < http://www.sapientia.pucsp.br/tde_busca/arquivo.php?codArquivo=6895 >. Acesso em: 15 mai. 2014.
- CENTROID LINKAGE (UPGMC). Disponível em: < <http://web.stanford.edu/~maureenh/quals/html/ml/node78.html> >. Acesso: 15.set.2013.
- CHALUPA, D., POSPICHAL, J., , 2012, “Metaheuristically Optimized Multicriteria Clustering for Medium-scale Networks”, *SOCO, Advances in Intelligent Systems and Computing*, v.188, pp. 333-346, Nova York, Springer,. Disponível em: <: <http://www2.fiit.stuba.sk/~chalupa/publ/soco12.pdf> >. Acesso em: 01 mai. 2013.
- CHARLES, *Spectral Clustering: A quick Overview*, 2012, Material online. Disponível em: <: <http://charlesmartin14.wordpress.com/2012/10/09/spectral-clustering/> >. Acesso em: 01 mai. 2013.
- CHEUNG, F., M., VIJVER, F., J., R., V., LEONG, F., T., L., 2011, “Toward a New Approach to the Study of Personality in Culture”, *American Psychologist*, v.66, n.7, (Out), pp. 593-603. Disponível em: < <http://www.apa.org/pubs/journals/releases/amp-66-7-593.pdf> >. Acesso em: 01 set. 2012.

- CHRETIENNE, Ph. Task “Schedulling over Distributed Memory Machines”, Workshop on Parallel and Distributed Algorithms, 1989, North Holland, pp. 165-176. In: GERASOULIS, A., YANG, T., 1992, pp.280-281.
- CHTIOUI, Y., PANIGRAHI, S., MARSH, R., 1998, “Conjugate Gradient and Approximate Newton Methods for an Optimal Probabilistic Neural Network for Food Color Classification”, *Optical Engineering -Bellingham- International Society for Optical Engineering*, v.37, n. 11 (Nov), pp. 3015-3023.
- CHUI, W., ZHOU, H., QU, H., WONG, P., C., LI, X., 2008, “Geometry-Based Edge Clustering for Graph Visualization” *IEEE. Transactions on Visualization and Computer Graphics*, v. 14, n. 6 (Dez), pp. 1277-1284.
- COLLAZO, R. A. , 2009, Aplicação de “Support Vector Machines” à Classificação do Risco de Morte de Pacientes com Síndrome Coronariana Aguda.Dissertação de de M. SC., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- COLEMAR A., 2013, *Esparsidade*, Apostila da Escola de Engenharia Elétrica, Mecânica e de Computação da Universidade Federal de Goiás, Rio de Janeiro, UFG. Disponível em: <: <http://www.eee.ufg.br/~colemar/Espars01.pdf> >. Acesso em: 03 mai. 2013.
- COTTRELL, M., HAMMER, B., HASENFUB, A., VILLMANN, T., 2005, “Batch Neural Gas”, *International Workshop on Self-Organizing Maps, WSOM 2005*, Paris. Disponível em: < <https://samos.univ-paris1.fr/archives/WSOM05/papers/WSOM2005-7.pdf> >. Acesso em: 01 mai. 2012.
- COSTA, P. T., MC CRAE, R., R., 1976, “Age Differences in Personality Structure: A Cluster Analytic Approach”, *Journal of Gerontology*, v.31, n.5, pp. 564-570. In: Big Five Personality, 2012.
- COSTA, P. T., MC CRAE, R., R., 1985, “The NEO Personality Inventory: Manual Form S and Form R”, Psychological Assessment Resources. PAR. In: Big Five Personality, 2012.
- COSTA, P. T., MC CRAE, R., R., 1992, “Revised NEO-PI-R Personality Inventory (NEO-PI-R) and NEO-PI-R Five-Factor Inventory (NEO-PIR-FFI) Professional Manual” In: CATTELL, H., E., P., MEAD, A., D., 2007, pp. 137-138.

- DEITEL, P., DEITEL, H., *Java Como Programar*, 8 ed., São Paulo, Pearson. Disponível em: <: http://www.4shared.com/get/UJ2GKc5d/livro_java_como_programar_edic.html >. Acesso em: 01 mai. 2012.
- DE RAAD, 2000, *The Big Five Personality Factors, The Psycholexical Approach to Personality*, First printing, Washington, Hofrege & Huber Publishers.
- DEMUTH, H., BEALE, M., 2001, *Neural Network Toolbox, Computation, Visualization, Programming, Matlab User's Guide*, Seventh printing – minor revisions, Massachusetts, The MathWorks, Inc.
- DIGMAN, J., M., “Five Trait Dimensions: Development, Stability, and Utility”, 1989, *Journal of Personality*, v.57, n.2, pp. 195-214. In: BIG FIVE, 2012.
- DONGEN, S. V., *Graph Clustering by Flow Simulation*, 2000, Tese de D. SC, University of Utrecht, Utrecht, Holanda. Disponível em: <: <http://micans.org/mcl/lit/svdthesis.pdf.gz> >. Acesso em: 01 mai. 2012.
- DRIVER, H. E., KROEBER, A. L., 1932, “Quantitative Expression of Cultural Relationships”, *University of California Publications in American Archaeology and Ethnology*, v.31 n.4, (Dez), pp. 211-256. Disponível em: <: <http://digitalassets.lib.berkeley.edu/anthpubs/ucb/text/ucp031-005.pdf> >. Acesso em: 01 mai. 2012.
- DUCH, W., ADAMCZAK, R., GRABEZEWSKI, K., 1999, “Neural Methods for Analysis of Psychometric Data”, *Proceedings of the 5th International Conference on Engineering Applications of Neural Networks (EANN'99)*, (Set), pp. 45-50. Disponível em: <: www.fizyka.umk.pl/publications/kmk/99eann.pdf >. Acesso em: 01 mai. 2012.
- ECLIPSE IDE FOR JAVA EE DEVELOPERS. Disponível em: < <https://www.eclipse.org> >. Acesso: 15.dez.2011.
- ESTER, M., KRIEGEL, H. P., SANDER, J., XU, X., 1996, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases With Noise”, *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, KDD-96*, AAAI Press, pp. 226-231. Disponível em: <: <http://www2.cs.uh.edu/~ceick/7363/Papers/dbscan.pdf> >. Acesso em: 23 fev. 2014.

- FISKE, D., W., 1949, “Consistency of the Factorial Structures of Personality Ratings From Different Sources, *Journal of Abnormal and Social Psychology*, v.1. n.44. In: DE RAAD, 2000, pp. 6-7.
- GAMMA, E., HELM, R., JOHNSON, R., VLISSIDES, J., *Design Patterns. Elements of Reusable Object-Oriented Software*, 1 ed., São Paulo, Boston, Professional Computing Series. Disponível em: < <http://www.uml.org.cn/c++/pdf/DesignPatterns.pdf> >. Acesso em: 14 de janeiro. 2012.
- GERASOULIS, A., YANG, T., 1992, “A comparison of clustering heuristics for scheduling directed acyclic graphs on multiprocessors”, *Journal of Parallel and Distributed Computing*, v.16 n.4, (Dez), pp. 276-291. Disponível em: <: <http://charm.cs.uiuc.edu/users/arya/docs/1.pdf> >. Acesso em: 01 mai. 2012.
- GOLDBERG, L., R., SPIELBERGER, C., D., BUTCHER, J., N., 1982, “From Ace to Zombie: Explorations in the Language of Personality”, *Advances in Personality Assessment*, v.1, Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey. Disponível em: < http://projects.ori.org/lrg/PDFs_papers/AcetoZombie.pdf >. Acesso em: 01 set. 2012.
- GOLDBERG, L., R., 1993, “The Structure of Phenotypic Personality Traits”, *American Psychologist*, v.48, n.1, pp.26-34. Disponível em: < http://projects.ori.org/lrg/PDFs_papers/Goldberg.Am.Psych.1993.pdf >. Acesso em: 01 set. 2012.
- GRAPHVIZ. Disponível em: <<http://www.graphviz.org/>>. Acesso: 15.jan.2014.
- HAGENBUCHNER, M., SPERDUTI, A., TSOI, A. C., 2003, “A self-organizing map for adaptive processing of structured data”, *IEEE Transactions on Neural Networks*, v.14 n.3, (Mai), pp. 491-505. Disponível em: <: <http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1145&context=infopapers> >. Acesso em: 01 mai. 2012.
- HAYKIN, S., 2005, *Neural Networks a Comprehensive Foundation*. 2.ed. Patparganj, Delhi, India, Prentice Hall.
- HETTEMA, J., DEARY, I. J., 1993, “Biological and social approaches to individuality: towards a common paradigm. In: HETTEMA, J, DEARY, I. J., “Foundations of

- personality, Dordrecht, The Netherlands: Kluwer In: MATTHEWS, I. J., WHITEMAN, M. C., 2003, p.6.
- HIGGINSON, B., PHELAN, P., 1986, "Word Cluster: A Strategy For Synonym Development", *Reading Horizons*, , v.26 Iss.3, article 4., (Abr), pp. 491-505. Disponível em: <: http://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=1827&context=reading_horizons >. Acesso em: 05 mar. 2012.
- HINNE, M., MARCHIORI, E., 2011, "Cutting Graphs Using Competing Ant Colonies and an Edge Clustering Heuristic", *EvoCOP of Lectures Notes in Computer Science*, v.6622, n.1, (), pp. 60-71 Disponível em: <: http://www.cs.ru.nl/~mhinne/papers/2011_Cutting_graphs_with_ants_and_edge_clustering_heuristic.pdf >. Acesso em: 03 mai. 2012.
- HINES, J. W., 1997, *Matlab Supplement to Fuzzy and Neural Approaches in Engineering*. 1 ed. Canada, John Wiley and Sons.
- JACKSON, D. N., PAUNONEN, S. V., 2000, "What is Beyond the Big Five? Plenty!", *Journal of Personality*, v.68, n.5, (Out), pp. 821-835. Disponível em: <: http://www.subjectpool.com/ed_teach/y4person/1_intro/refs/whatsbeyondthebig-5.pdf >. Acesso em: 01 mai. 2012.
- JANG, K., L., LIVESLEY, W., J., 1996, "Heritability of The Big Five Personality Dimensions and Their Facets: a Twin Study", *Journal of Personality*, v.64, n.3, (Set), pp. 577-592. Disponível em: < www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CCYQFjAB&url=http%3A%2F%2Fwww.researchgate.net%2Fpublication%2F14425701_Heritability_of_the_big_five_personality_dimensions_and_their_facets_a_twin_study%2Flinks%2F0c960525c3b9b5cde0000000&ei=u5XfU6C5FM3gsATP8oGgDg&usg=AFQjCNE5gw0qYERaT27eoLzjmHINa7M5Ng&bvm=bv.72197243,d.b2U >. Acesso em: 01 set. 2012.
- JOB INTERVIEW STRENGTH AND WEAKNESS. FORMULÁRIO COM PERGUNTAS SOBRE PONTOS FORTES E FRACOS PARA CANDIDATOS DE UM PROCESSO SELETIVO. Disponível em: < <http://www.collegegrad.com/interview/jobinterviewstrengthandweakness.shtml> >. Acesso em: 01 set 2013.

- KARSON, S., O'DELL, J. W., 1976, A Guide to the Clinical Use of the 16PF, IL: Institute for Personality & Ability Testing, Champaign. In: Big Five Personality Traits, 2012. In: Big Five Personality Traits, 2012.
- KIM, S. J., BROWNE, J. C., “A General Approach to Mapping of Parallel Computation Upon Multiprocessor Architectures”, *International Conference of Parallel Processing*, 1988, (Abr), v.3, n.1, pp. 1-8. In: GERASOULIS, A., YANG, T., 1992, pp.280-281.
- KLINKOSZ, W., SEKOWSKI, A., E., 2008, “National Taxonomies, Adjective Makers and Inventories: Three Directions of Application of Lexical Approach to Personality”, *Roczniki Psychologiczne (Annals of Psychology) (1507-7888)* XI, v.1, n.11, pp.152-160. Disponível em: < http://www.kul.pl/files/1024/Roczniki_Psychologiczne/2008/1/RPsych_2008nr1_152-160_Mlacic.pdf >. Acesso em: 01 set.2012.
- KOFI, A., HUNTER, A., MENG, H., YUE, S., HOB DEN, M., PRIESTLEY, N., HOB DEN, P., PETTIT, C., “A Binary Self-Organizing Map and its FPGA Implementation”, *Proceedings of International Joint Conference on Neural Networks, 2009*, (Jun), pp. 14-19.
- KOHONEN, T., 2001, *Self-Organizing Maps*, 3 ed, New York, Springer.
- KRUG, S., E., 1986, “A Large Scale Cross-Validation of Second-Order Personality Structure Defined by the 16PF”, *Psychological reports*, v,59, n.2, pp. 683-693. In: Big Five Personality Traits, 2012.
- KRUG, S., E., 1989, “A The 16PF Tradition in Contemporary Personality Assessment”, *97th Meeting of the American Psychological Association*, APA, New Orleans, 1989, (Aug), pp.1-11. Disponível em: < <http://files.eric.ed.gov/fulltext/ED316561.pdf> >. Acesso em: 15 set 2013.
- KUN-SEOK, O., FENG, Y., KANEKO, K., FUKUOKA H., H., KWANGJU, “SOM-Based R*-Tree for Similarity Retrieval”, *Proceedings Seventh International Conference on Database Systems for Advanced Applications. DASFAA 2001*, (Abr), pp. 182-189. Disponível em: < http://www.comp.nus.edu.sg/~lingtw/dasfaa_proceedings/dasfaa2001/00916377.pdf >. Acesso em: 23 fev. 2014.
- LINS, M., P., E., DINIZ, M., E., 2012, *Percepção e Estruturação de Problemas Sociais Utilizando Mapas Cognitivos*, Associação Brasileira de Engenharia de Produção,

- Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-65132012000100010 >. Acesso em: 15 mai. 2014.
- MACROPOL, K., *Clustering on Graphs: The Markov Cluster Algorithm (MCL)*, University of California, Santa Barbara, Department of Computer Science, 2009. Disponível em: <: http://www.cs.ucsb.edu/~xyan/classes/CS595D-2009winter/MCL_Presentation2.pdf >. Acesso em: 01 mai. 2012.
- MATSUO, Y., UCHIYAMA, K., SAKAKI, T., ISHIZUKA M., 2006, “Graph-Based Word Clustering using a Web Search Engine”, *Proceedings of The 2006 Conference on Empirical Methods in Natural Language Processing*, páginas 542-550, Sydney. Disponível em: <: <http://www.miv.t.u-tokyo.ac.jp/papers/matsuo-sakaki-EMNL06.pdf> >. Acesso em: 01 ago. 2013.
- MATTHEWS, I. J., WHITEMAN, M. C., 2003, *Personality Traits*. 2 ed. Cambridge, Cambridge University Press. Disponível em: <: <http://elib.fk.uwks.ac.id/asset/archieve/e-book/PSYCHIATRIC-%20ILMU%20PENYAKIT%20JIWA/Personality%20Traits,%202nd%20Ed.pdf> >. Acesso em: 01 mar. 2011.
- MCCRAE, R., R., COSTA, P. T., 1987, “Validation of the Five-Factor Model of Personality Across Instruments and Observers”, *Journal of Personality and Social Psychology*, v.1, n.52. pp. 81-90. In: Big Five Personality, 2012.
- MCCRAE, R., R., COSTA, P. T., OSTENDORF, F., ANGLEITNER, A., HREBÍCKOVÁ, M., AVIA, M.D., SANZ, J., SÁNCHEZ-BERNARDOS, M., L., KUSDIL, M. E., WOODFIELD, R., SAUNDERS, P. R., SMITH, P., B., 2000, “Nature Over Nurture: Temperament, Personality and Life Span Development”, *Journal of Personality and Social Psychology*, v.78, n.1, (Out), pp. 173-186. Disponível em: <: http://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CCgQFjAB&url=http%3A%2F%2Fwww.researchgate.net%2Fprofile%2FPaul_Costa3%2Fpublication%2F12660310_Nature_over_nurture_temperament_personality_and_life_span_development%2Flinks%2F00b49515b9ad9066e1000000&ei=A73aU_DjIdOsyATK_oDwBg&usg=AFQjCNGsVIx7Djynl4DIy9wcU7vboq32WQ&bvm=bv.72185853,d.aWw&cad=rja >. Acesso em: 01 mai. 2013.

- MCCRAE, R., R., OLIVER, J., P., “An Introduction to the Five-Factor Model and its Applications”, 1992, *Journal of Personality*, v.2, n.60, pp. 175-215. Disponível em: < http://psych.colorado.edu/~carey/courses/psyc5112/readings/psnbig5_mccrae03.pdf >. Acesso em: 01 set. 2012.
- MCL. Disponível em: < <http://micans.org/mcl/> >. Acesso em: 01-julho.2013.
- MCL (MARKOV CLUSTERING ALGORITHM) código em Java. Disponível em: < <http://sourceforge.net/directory/os:windows/freshness:recently-updated/?q=MCL> >. Acesso: 15.mai.2013.
- MYSQL. Disponível em: < <http://www.mysql.com/> >. Acesso em: 01.janeiro.2012.
- NETBEANS. Disponível em: < <https://netbeans.org/> >. Acesso em: 01.janeiro.2012.
- NOACK, A., “Energy Models for Graph Clustering”, 2007, *Journal of Graph Algorithms and Applications*, v.11, n.2, pp.453-480. Disponível em: < ftp://garryowen.csisdmsz.ul.ie/pub/healyp/public_html/jgaa-gd05/Noack/Noack_Clustering.pdf >. Acesso em: 15.mai.2013.
- NORMAN, W. T., 1963, “Toward an Adequate Taxonomy of Personality Attributes: Replicated Factor Structure in Peer Nomination Personality Ratings”, *The Journal of Abnormal and Social Psychology*, v.66, n.6, (Jun), pp. 574-583. In: MCCRAE, R., R., OLIVER, J., P., 1992, pp. 177-180.
- NORMAN, W., T., GOLDBERG, L. R., 1966, “Raters, Ratees, and Randomness in Personality Structure”, *Journal of Personality and Social Psychology*, v.4, n.6, pp.681-691. Disponível em: < http://projects.ori.org/lrg/PDFs_papers/raters.ratees.norman.pdf >. Acesso em: 01.set.2012.
- OSTROSKI, A., MENONCINI, L., 2009, “Teoria dos Grafos e Aplicações”, *Synerismus scyentifica UTFPR – 13º Encontro Regional de Matemática Aplicada e Computacional – XIII ERMAC*, v.2, n.4, (Out), pp. 1-6.
- PAPADIMITRIOU, C., YANNAKAKIS, M., Task “Towards an Architecture-Independent Analysis of Parallel Algorithms”, *SIAM Journal of Computing*, 1989, vol.19, Nova York, pp. 322-328. In: GERASOULIS, A., YANG, T., 1992, pp.280-281.
- PARVIN, H., MINAEI-BIDGOLI, B., DAHBASHI, A., “Improving Persian Text Classification Using Persian Thesaurus”, 2012, *CIARP'11 Proceedings of the 16th Iberoamerican Congress conference on Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, v.1, n.1, Berlim, Springer-Verlag

- Berlin Heidelberg, pp. 391-398. Disponível em: <:
<http://202.154.59.182/mfile/files/Information%20System/Progress%20in%20Pattern%20Recognition,%20Image%20Analysis,%20Computer%20Vision,%20and%20Applications%3B%2016th%20CIARP%202011/Chapter%2046%20Improving%20Persian%20Text%20Classification%20Using%20Persian%20Thesaurus.pdf> >. Acesso em: 01 mai. 2013.
- PATTERSON, D. W., 1996, *Artificial Neural Networks, Theory and Applications*, 1 ed., New York, Prentice Hall
- PASSOS, A. R. F., 2010, Aplicação de Redes Neurais Probabilísticas à Classificação do Risco de Morte de Pacientes com Síndrome Coronariana Aguda. Dissertação de de M. SC., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- PEABODY, D., GOLDBERG, L., R., 1989, “Some Determinants of Factor Structures From Personality-Trait Descriptors”, *Journal of Personality and Social Psychology*, v.57, n.3, pp. 552-567. Disponível em: <
http://projects.ori.org/lrg/PDFs_papers/PeabodyGoldberg89JPSP.pdf >. Acesso em: 01 set. 2012.
- PEREIRA, B. B., RAO, C. R., 2009, *Data Mining Using Neural Networks: A Guide for Statisticians*, Disponível em:
 <http://textbookrevolution.org/index.php/Data_Mining_using_Neural_Networks:A_Guide_for_Staticians >. Acesso em: 15 de dezembro. 2011..
- PHP. Disponível em: < <http://php.net/> >. Acesso em: 12. abril.2011.
- PHP. COMPILAÇÃO 64 BITS. Disponível em: < <http://www.anindya.com/> >. Acesso em: 23. abril.2011.
- RAYMOND, T. N., JIAWEI, H., 1994, “Efficient and Effective Clustering Methods for Spatial Data Mining”, *Proceedings of the 20th VLDB Conference*, Santiago, Chile, pp. 144-155. Disponível em: <www.vldb.org/conf/1994/P144.PDF >. Acesso em: 15 dez. 2013.
- REIS. M., M., *Processos Estocásticos*, 2014, Apostila da UFSC – Universidade Federal de Santa Catarina. Disponível em: < <http://www.inf.ufsc.br/~marcelo/processos1.pdf> >. Acesso em: 15 mai. 2014.

- RIPLEY, B. D., 2004, *Pattern Recognition and Neural Networks*. 7ed. New York, Cambridge University Press.
- ROTHMANN, S., COETZER, E., P., 2003, “The Big Five Personality Dimensions and Job Performance”, *SA Journal of Industrial Psychology*, v.29, n.1, pp.68-74. Disponível em: < <http://www.sajip.co.za/index.php/sajip/article/viewFile/88/84> >. Acesso em: 01 set. 2012.
- ROGET, P., M., ROGET’S THESAURUS. In: THESAURUS, 2013.
- SANSONNET, J., P., BOUCHET, F., 2010, “Extraction of Agent Psychological Behaviors from Glosses of WordNet Personality Adjectives”, *Proceedings of the 8th European Workshop on Multi-Agent Systems (EUMAS’10)*, Paris, França, pp 1- 15. Disponível em:
< http://perso.limsi.fr/jps/research/rmb/doc/10.A_EUMAS.17dec10.pdf >. Acesso em: 15 de setembro. 2013.
- SARKAR, V., 1989, *Partitioning and Scheduling Parallel Programs for Multiprocessors*. 1ed. Cambridge, The MIT Press. Disponível em: <: www.cs.rice.edu/~vs3/PDF/Sarkar89.pdf >. Acesso em: 01 set. 2013.
- SAUCIER, G., GOLDBERG, L. R., 1996, “The Language of Personality: Lexical Perspectives on the Five-Factor Model”, In: WIGGINS, J., S., 1996, pp. 21-50. Disponível em: < http://projects.ori.org/lrg/PDFs_papers/Big.Five.Wiggins.Chapter.pdf >. Acesso em: 01 set. 2012.
- SCHAEFFER, S., E., 2007, “Graph Clustering”, *Computer Science Review*, v.1, pp. 27-64.
- SCHEINERMAN, E. R. 2008, *Matgraph: a Matlab Toolbox for Graph Theory*. Disponível em: <: <http://www.ams.jhu.edu/~ers/matgraph/matgraph.pdf> >. Acesso em: 07 set. 2011.
- SCHEINERMAN, E. R. 2008, *Matgraph Toolbox for Matlab*. Disponível em: <: <http://www.ams.jhu.edu/~ers/matgraph/matgraph.pdf> >. Acesso em: 07 set. 2011.
- SELIM, M., SIDDIK, S., GIAS, A., U., WADUD, A., A., KHALED, S., M., “A Genetic Algorithm for Software Design Migration from Structured to Object Oriented Paradigm”, 2014, 8th International Conference on Computer Engineering and

- Application, CEA 2014, Tenerife, Spain, v.1, n.1, (jan), pp. 187-192. Disponível em: <http://www.wseas.us/e-library/conferences/2014/Tenerife/INFORM/INFO_RM-26.pdf>. Acesso em: 21 out. 2014.
- SIERRA, . K., BATES, B., *Use a Cabeça! Java*, 2 ed., Rio de Janeiro, Alta Books. Disponível em: < http://www.4shared.com/get/c8uYajam/Use_a_Cabeca_Java.html >. Acesso em: 14 mar. 2011.
- SILVEIRA, D., S., MELO, V., A., NETTO, P., O., B., “AlgoDeGrafos: An Application to Assist in Course Lectures on Graph Theory”, 2009, *CLEI Electronic Journal*, v.12, n.1, (Abr), Paper 2.
- SINGH, J., K., MISRA, G., 2010, “A Psycholexical Study of Personality Trait Structure of Hindi Speaking Indians”, Dynamics the 2010 Congress of The International Association for Cross-Cultural Psychology, Melbourne, Australia. Disponível em: < http://www.indigenoupsych.org/News/An_Indian_Personality_Measure.pdf >. Acesso em: 01 set. 2012.
- SOUZA, C. R., PIZZOLATO, E. B., 2013, “Sign Language Recognition with Support Vector Machines and Hidden Conditional Random Fields”, *Lecture Notes in Computer Science, LNAI*, 7988, pp. 84-98. Disponível em: <: <http://www2.dc.ufscar.br/~cesar.souza/publications/mldm2013-csouza.pdf> >. Acesso em: 23 fev. 2014.
- SEDGEWICK, R., WAYNE K, 2010, *Ford Fulkerson*. Disponível em: <: <http://algs4.cs.princeton.edu/64maxflow/FordFulkerson.java.html> >. Acesso em: 07 set. 2011.
- SQL MANAGER. Disponível em: < <http://www.sqlmanager.net/> >. Acesso em: 12. abril.2011.
- THE FREE DICTIONARY. Disponível em: < <http://www.thefreedictionary.com/> >. Acesso em: 10.dezembro.2013.
- THESAURUS. Disponível em: < <http://www.thesaurus.com> >. Acesso em: 26.agosto.2013.
- THESAURUS. Disponível em: < http://en.wikipedia.org/wiki/Thesaurus#cite_note-Roget-1 >. Acesso em: 15.maio.2013.
- THURSTONE, L., L., 1947, *Multiple Factor Analysis*, Chicago University Press, Chicago, IL. Disponível em: < <http://stats.org.uk/factor-analysis/Thurstone1931.pdf> >. Acesso em: 15 de dezembro. 2013..

- TRADUTOR BABYLON. Disponível em: < <http://tradutor.babylon.com/ingles/portugues/> >. Acesso em: 26.agosto.2013.
- TRYON, R. C., BAILEY, D. E., 1970, *Cluster Analysis*. 1 ed., New York, McGrall-Hill.
- TUPES, E., C., CHRISTAL, R., E., 1961, “Recurrent Personality Factors on Trait Ratings”, *Personnel Laboratory Aeronautical Systems Division*, Air Force Systems Command, Laceland, Texas. Disponível em: < <http://www.dtic.mil/dtic/tr/fulltext/u2/267778.pdf> >. Acesso em: 01 set. 2012.
- VESANTO, J., ALHONIEMI, E., 2000, “Clustering of the Self-Organizing Map”, *IEEE Transactions os Neural Networks*, v.11 n.3, (Mai), pp. 586-600. Disponível em: <: <http://www.smartquant.com/references/NeuralNetworks/neural2.pdf> >. Acesso em: 01 mai. 2012.
- WALSH, T., 2003, *Distributional Clustering of Words for Text Classification*, Apostila da Rutgers University em *Learned representation in AI*, Nova Jersey. Disponível em: <: <https://www.cs.rutgers.edu/~mlittman/courses/lightai03/twalsh.ppt> >. Acesso em: 03 mai. 2013.
- WASSERMAN, L., 2004, *All of Statistics: a Concise Course in Statistical Inference*. 2 ed. New York, Springer.
- WASSERMAN, P., 1993, *Advanced Methods in Neural Computing*. 1 ed, New York, Van Nostrand Reinhold.
- WEBSTERS ONLINE. Disponível em: < <http://www.websters-online-dictionary.org/> >. Acesso: 17.01.2012.
- WONG, E. S. H., YOUNG, E. F., Y., MAK, W. K., 2003, “Clustering Based Acyclic Multi-way Partitioning”, *GLSVLSI Proceedings of The 13th ACM Great Lakes Symposium*, páginas 203-206, Nova York. Disponível em: <: <http://dl.acm.org/citation.cfm?id=764860> >. Acesso em: 01 set. 2013.
- WORD REFERENCE. Disponível em: < <http://www.wordreference.com/enpt/> >. Acesso: 26.08.2013.
- YONG, G., PEARCE, S., 2013, “A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis”, *Tutorials in Quantitative Methods for Psychology*, v.9,

n.2, (Jan), pp. 79-94. Disponível em: < <http://www.tqmp.org/Content/vol09-2/p079/p079.pdf> >. Acesso em: 01 set. 2013.

ZUBIN, T., 1938, “A Technique for Measuring Like-Mindedness”, *Journal of Abnormal and Social Psychology*, v. 33 n.1, (Out), pp. 508-516. Disponível em: <: <http://www.wpic.pitt.edu/research/biometrics/Publications/Biometrics%20Archives%20PDF/014-1938%20Zubin.pdf> >. Acesso em: 01 set. 2013.

ZUCKERMAN, EDWARD L., 2010, *Clinician's Thesaurus*. 7 ed. Belmont, The Guilford Press.

APÊNDICE

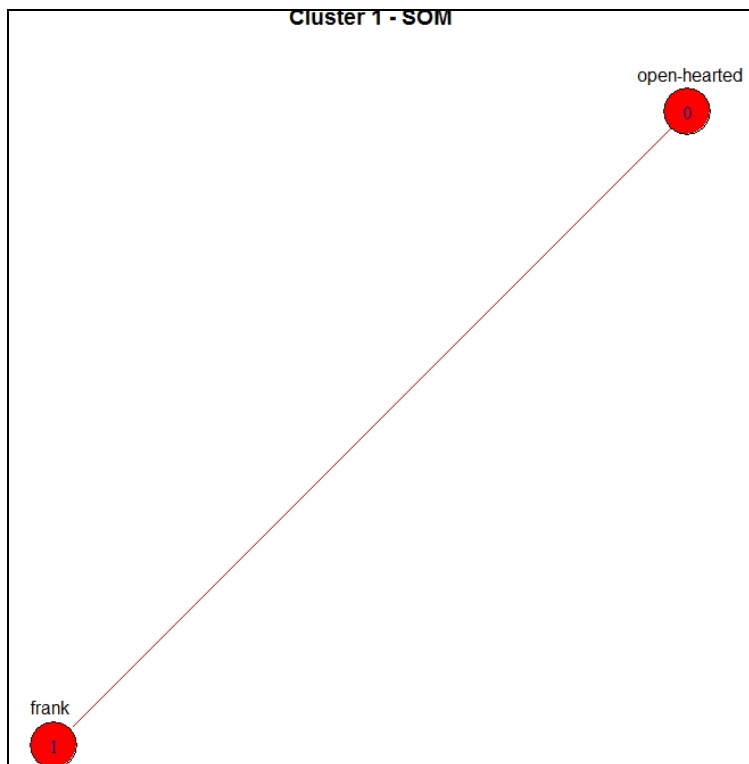


Figura 36: *Cluster 1* – Metodologia SOM

Nota: a rotulação dos vértices do cluster na figura não possui correlação com a rotulação apresentada em relatórios anteriores, tendo sido gerada inevitavelmente pelo código do *software* R. Não possui valor nenhum para fins de análise, devendo ser ignorada.

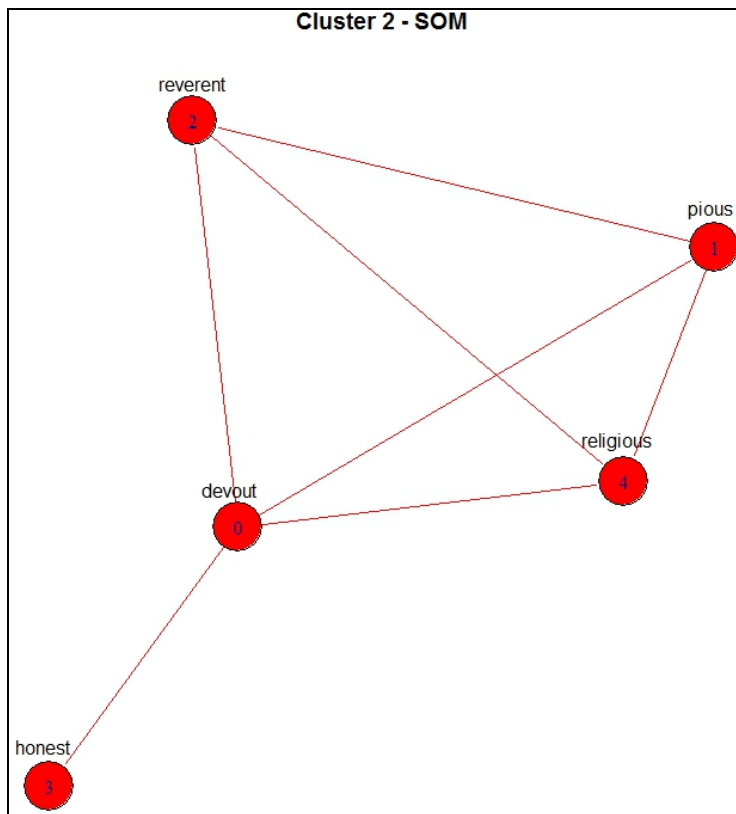


Figura 37: Cluster 2 – Metodologia SOM

Nota: idem da nota da figura 1.

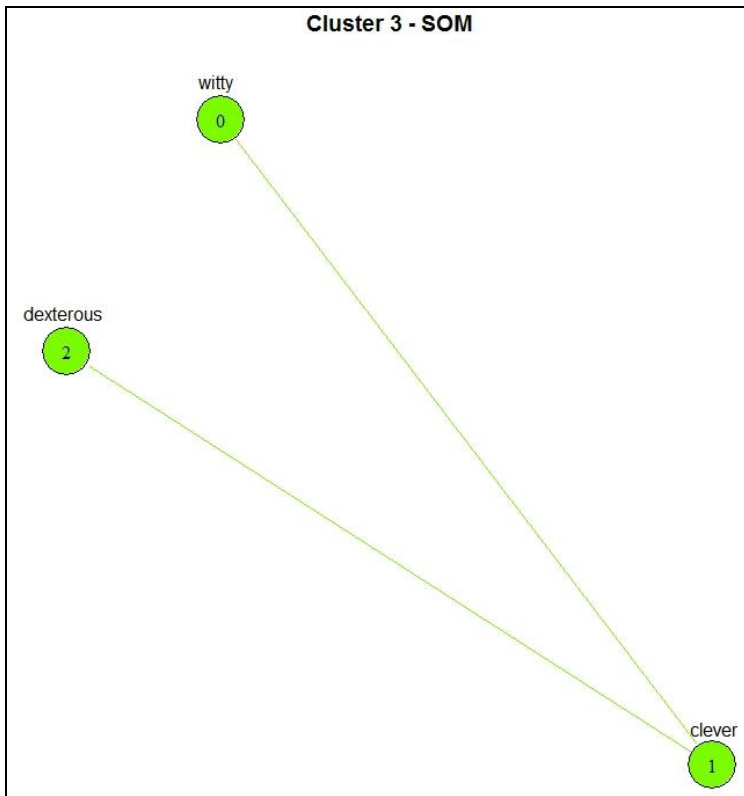


Figura 38: Cluster 3 – Metodologia SOM

Nota: idem da nota da figura 1.

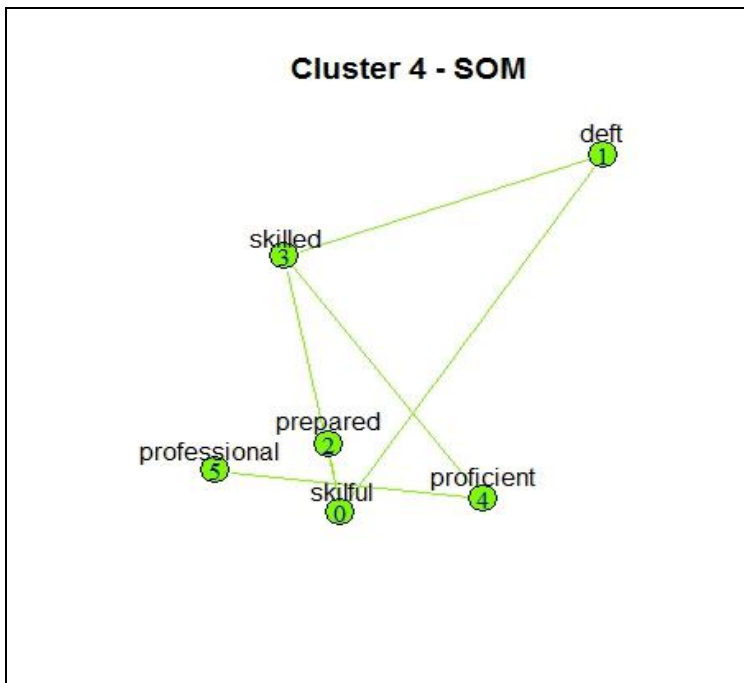


Figura 39: Cluster 4 – Metodologia SOM

Nota: idem da nota da figura 1.

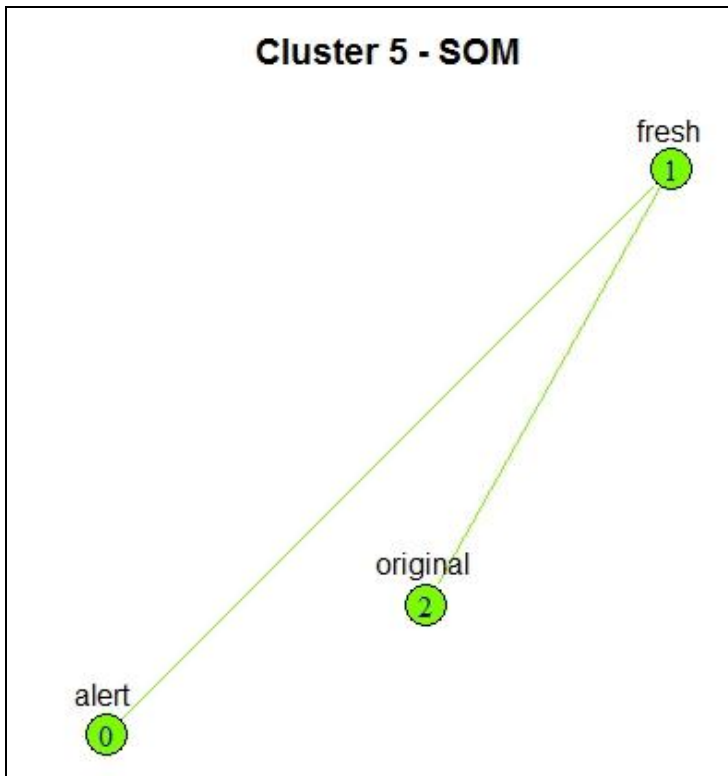


Figura 40: Cluster 5 – Metodologia SOM

Nota: idem da nota da figura 1.

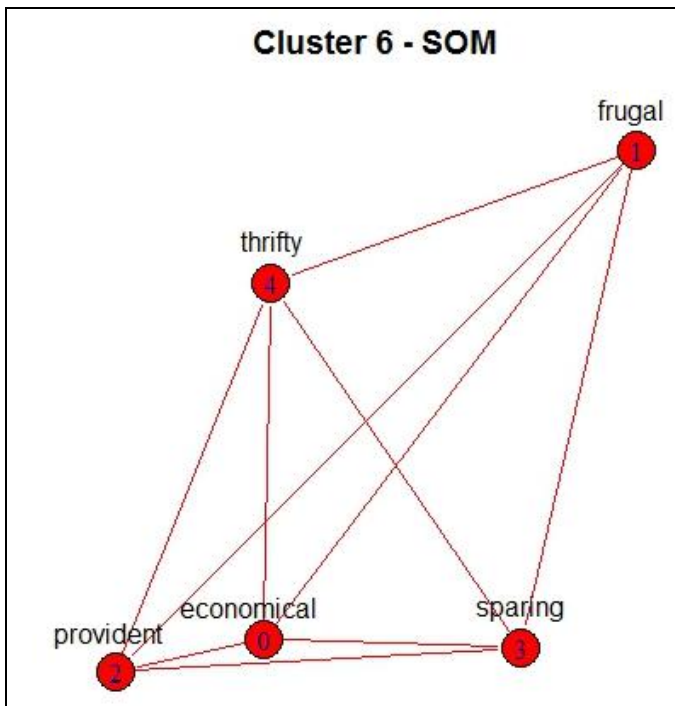


Figura 41: Cluster 6 – Metodologia SOM

Nota: idem da nota da figura 1.

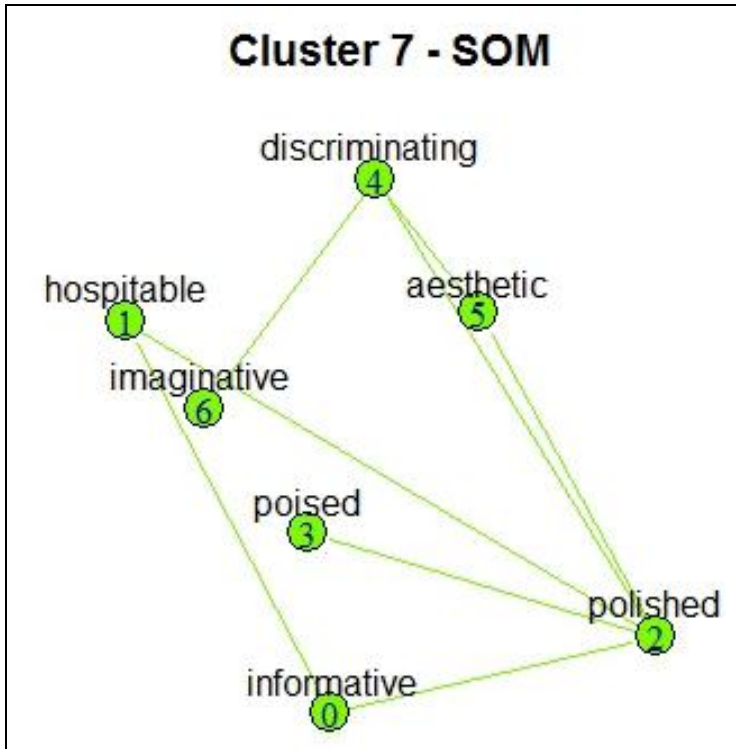


Figura 42: Cluster 7 – Metodologia SOM

Nota: idem da nota da figura 1.

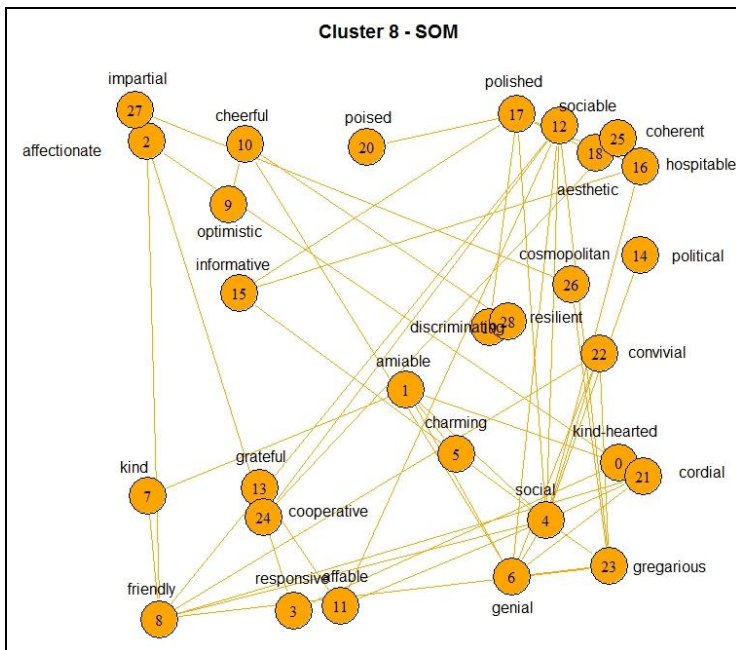


Figura 43: Cluster 8 – Metodologia SOM

Nota: idem da nota da figura 1.

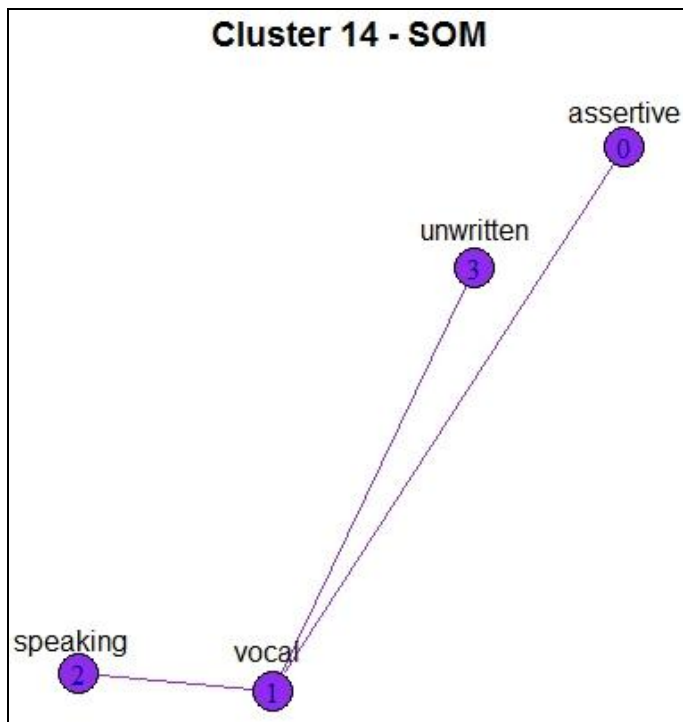


Figura 44: Cluster 14 – Metodologia SOM

Nota: idem da nota da figura 1.

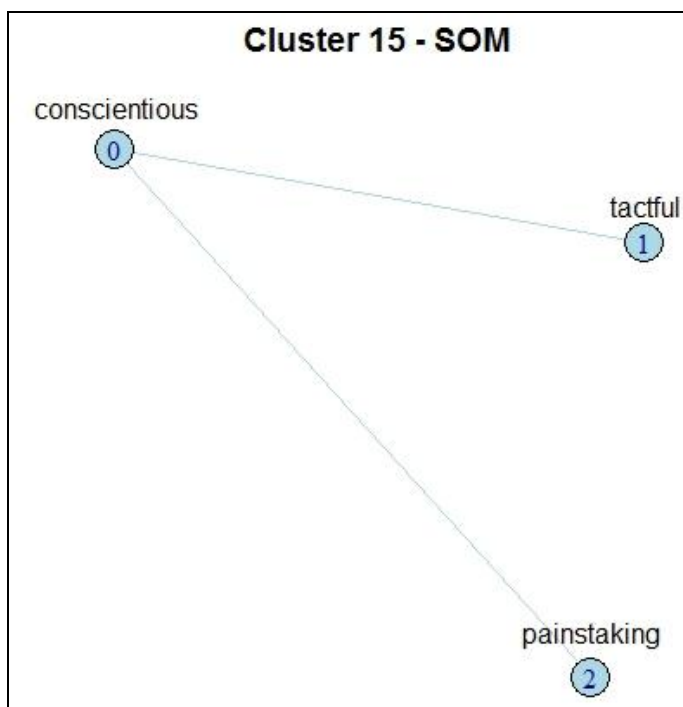


Figura 45: Cluster 15 – Metodologia SOM

Nota: idem da nota da figura 1.

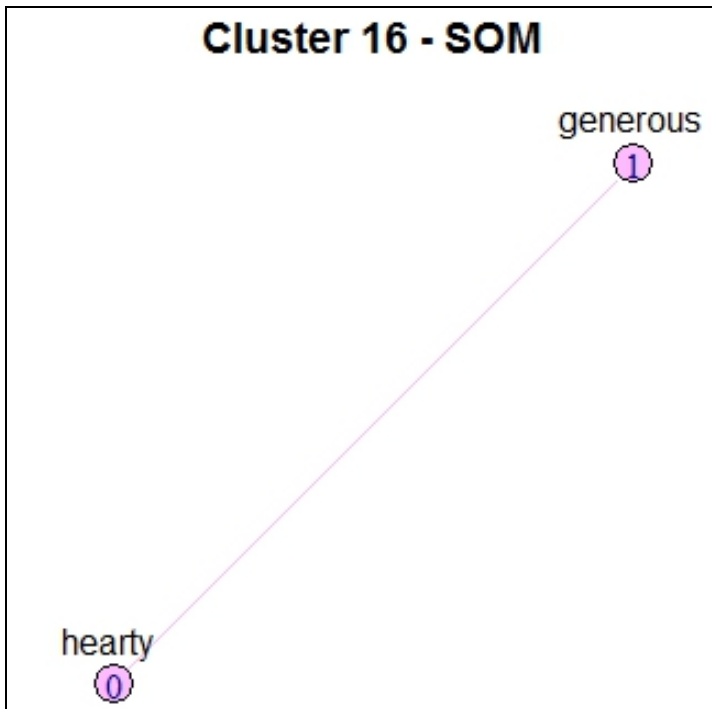


Figura 46: *Cluster 16*– Metodologia SOM

Nota: idem da nota da figura 1.

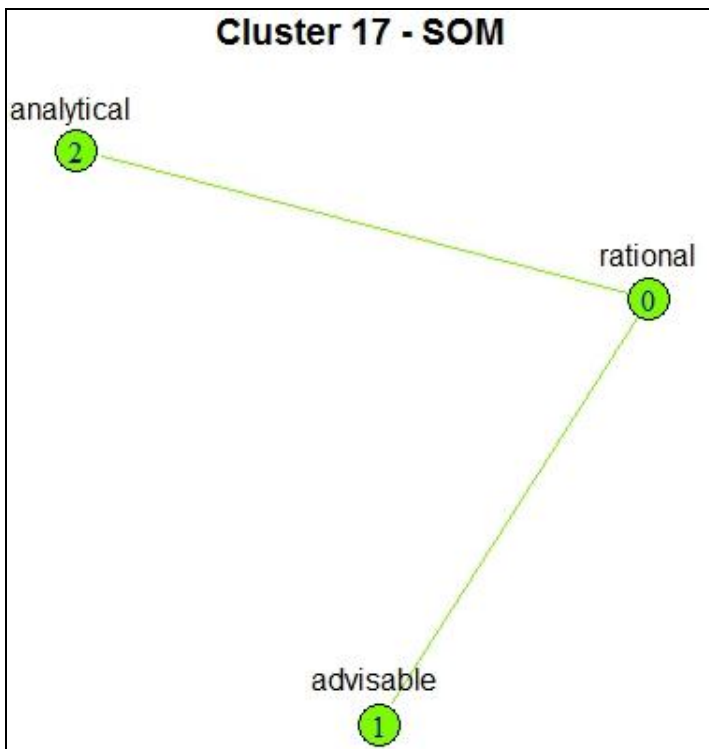


Figura 47: *Cluster 17* – Metodologia SOM

Nota: idem da nota da figura 1.

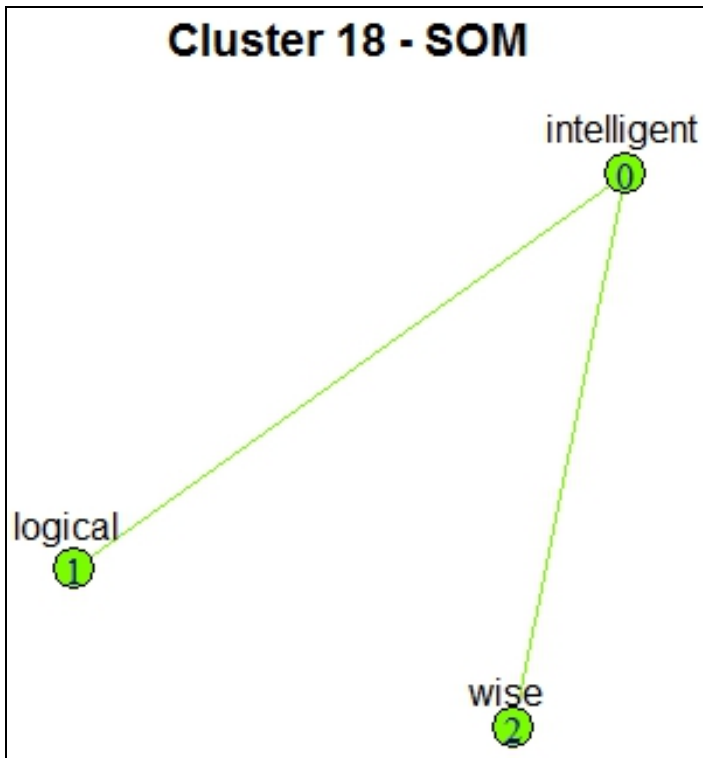


Figura 48: *Cluster 18* – Metodologia SOM

Nota: idem da nota da figura 1.

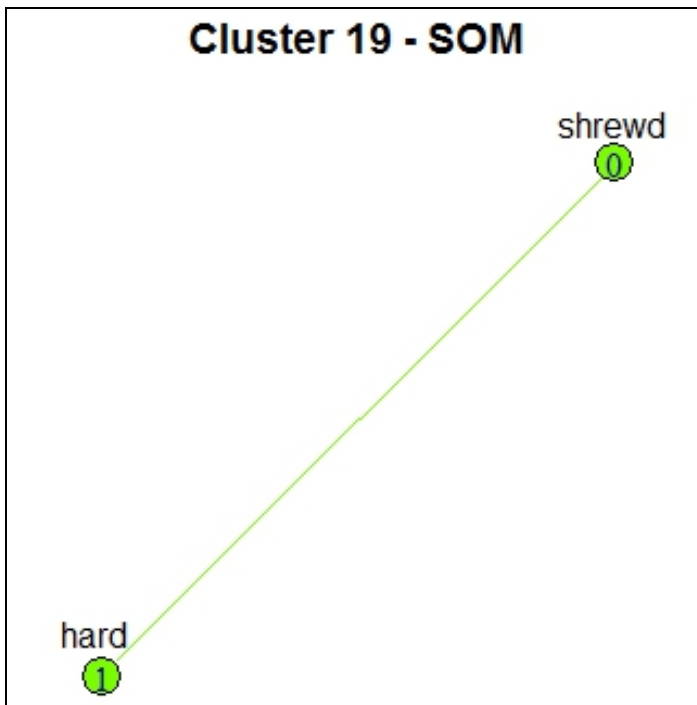


Figura 49: *Cluster 19* – Metodologia SOM

Nota: idem da nota da figura 1.

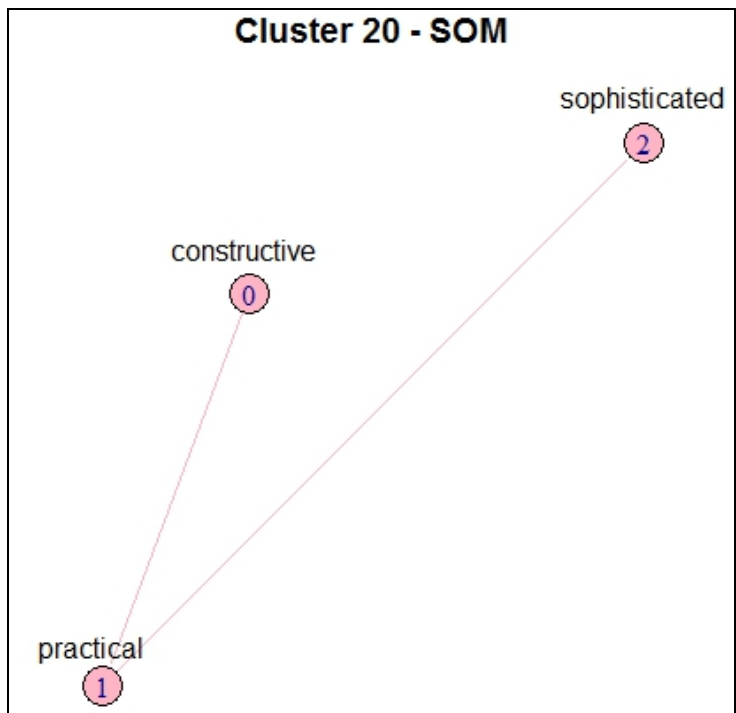


Figura 50: *Cluster 20* – Metodologia SOM

Nota: idem da nota da figura 1.

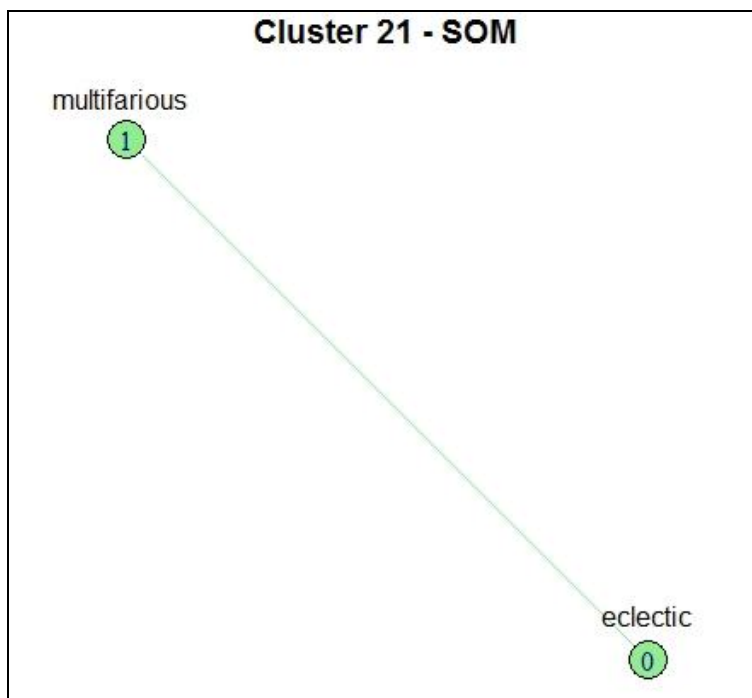


Figura 51: *Cluster 21* – Metodologia SOM

Nota: idem da nota da figura 1.

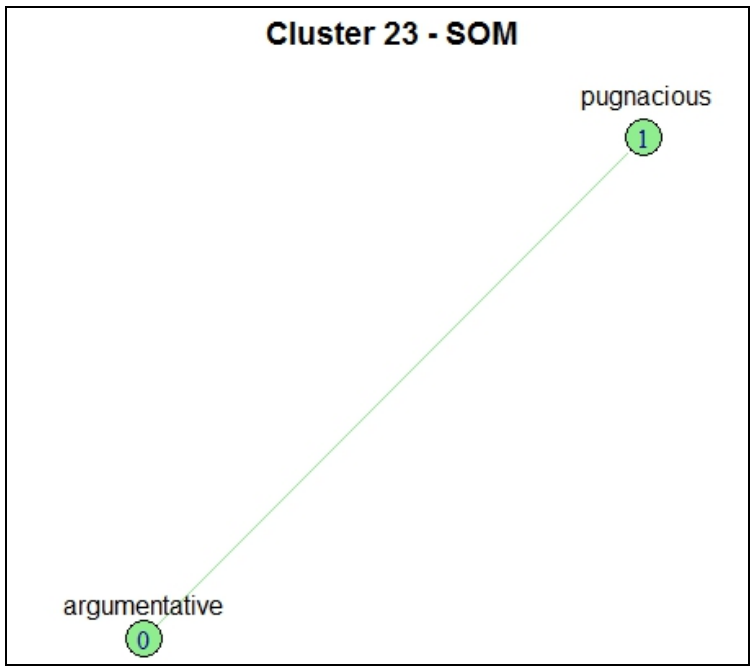


Figura 52 Cluster 23 – Metodologia SOM

Nota: idem da nota da figura 1.

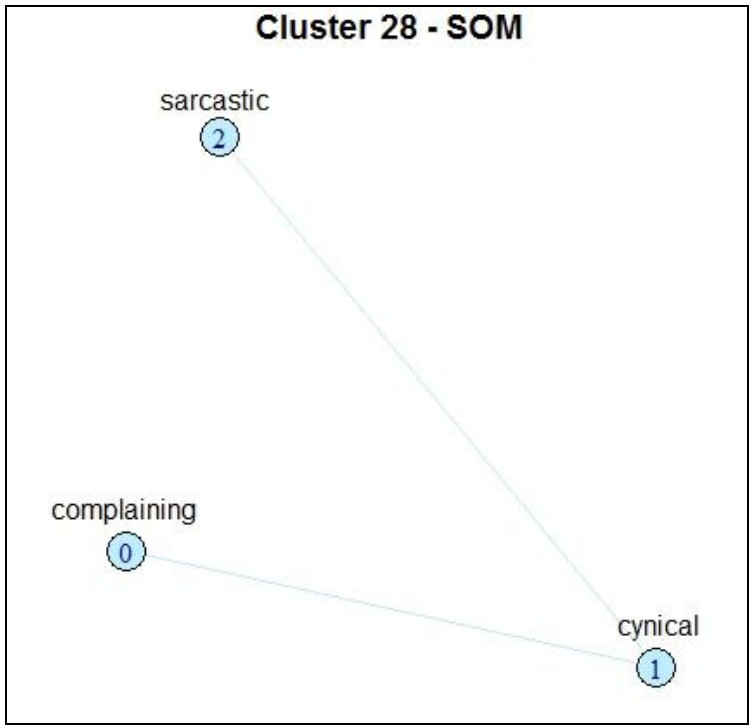


Figura 53: Cluster 28 – Metodologia SOM

Nota: idem da nota da figura 1.

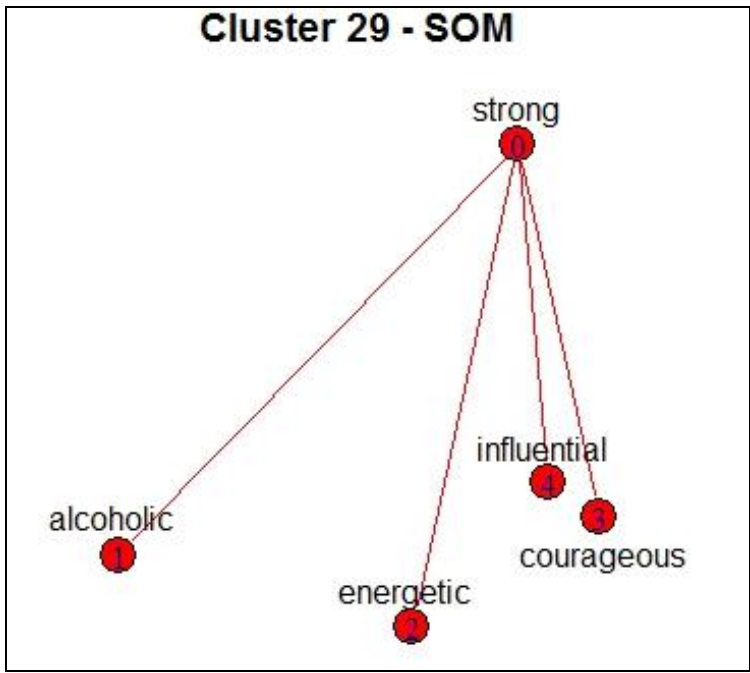


Figura 54: Cluster 29 – Metodologia SOM

Nota: idem da nota da figura 1.

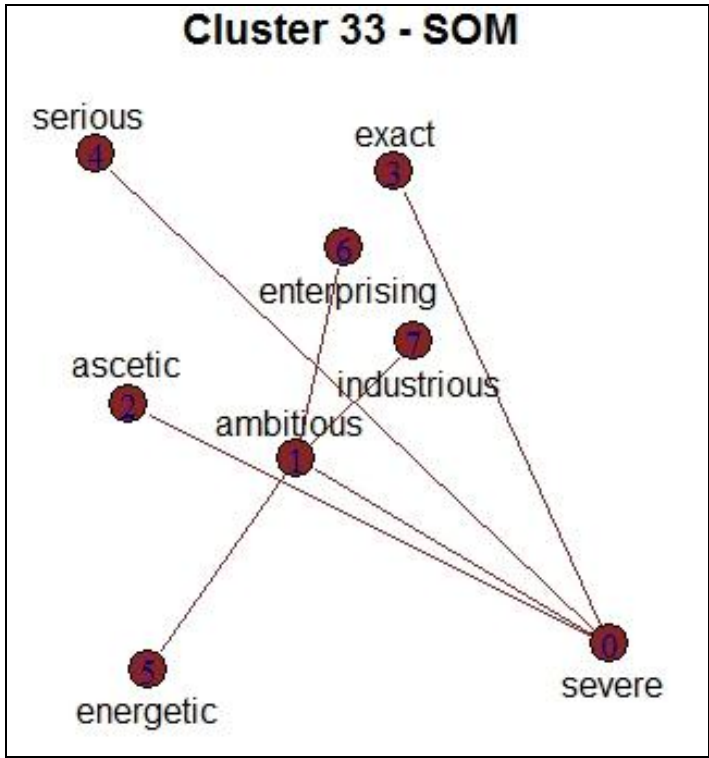


Figura 55: Cluster 33 – Metodologia SOM

Nota: idem da nota da figura 1.

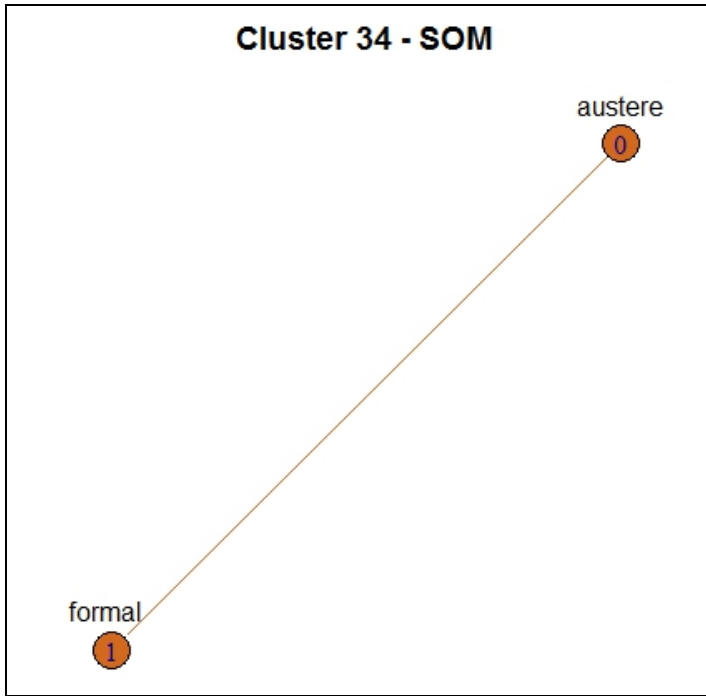


Figura 56: *Cluster 34* – Metodologia SOM

Nota: idem da nota da figura 1.

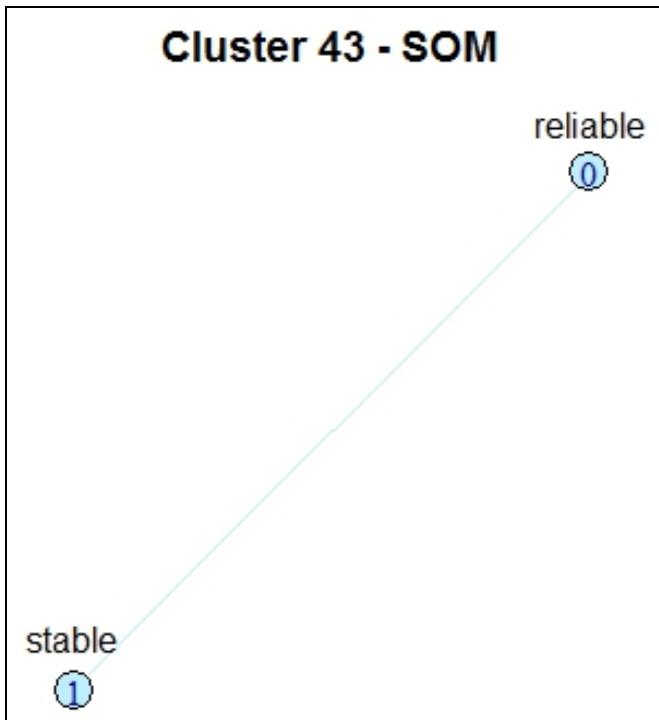


Figura 57: *Cluster 43* – Metodologia SOM

Nota: idem da nota da figura 1.

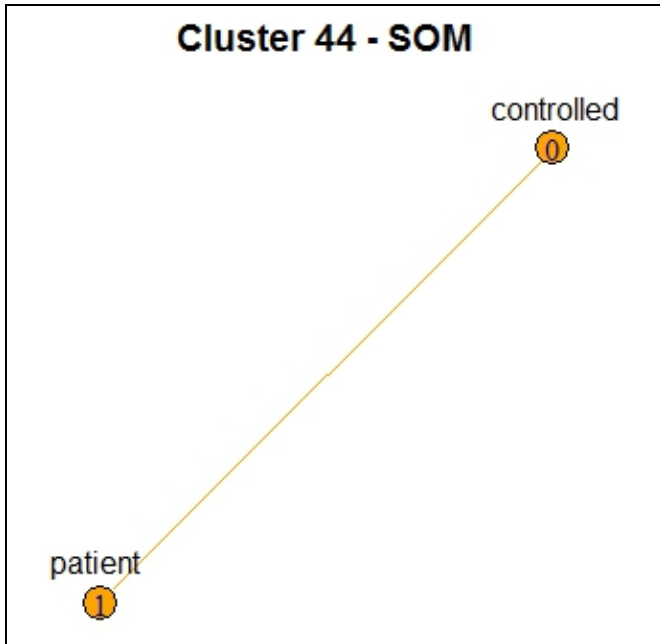


Figura 58: *Cluster 44* – Metodologia SOM

Nota: idem da nota da figura 1.

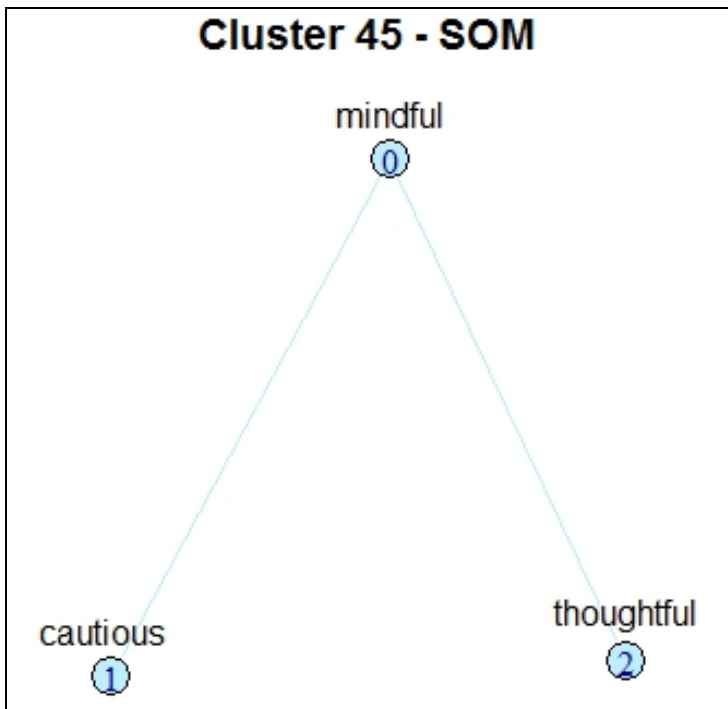


Figura 59: *Cluster 45* – Metodologia SOM

Nota: idem da nota da figura 1.

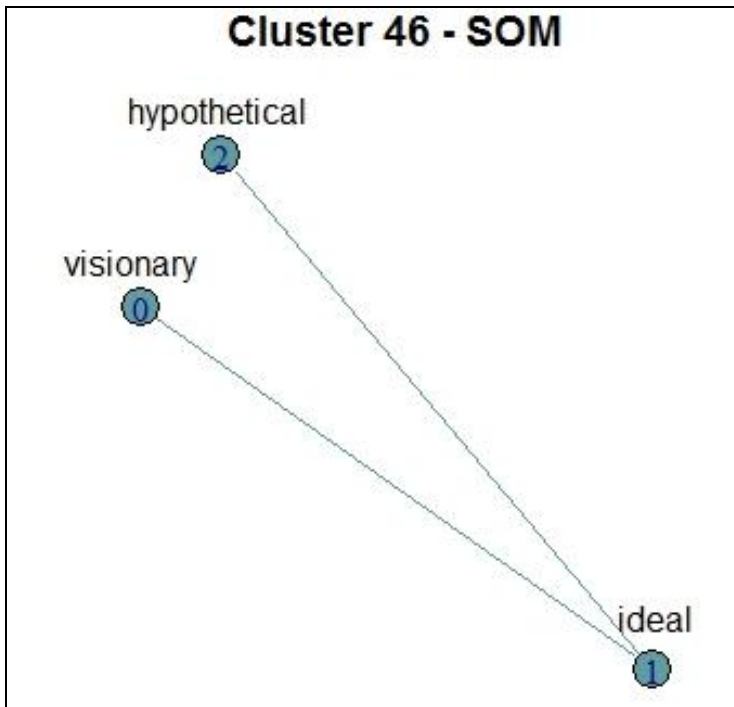


Figura 60: *Cluster 46* – Metodologia SOM

Nota: idem da nota da figura 1.

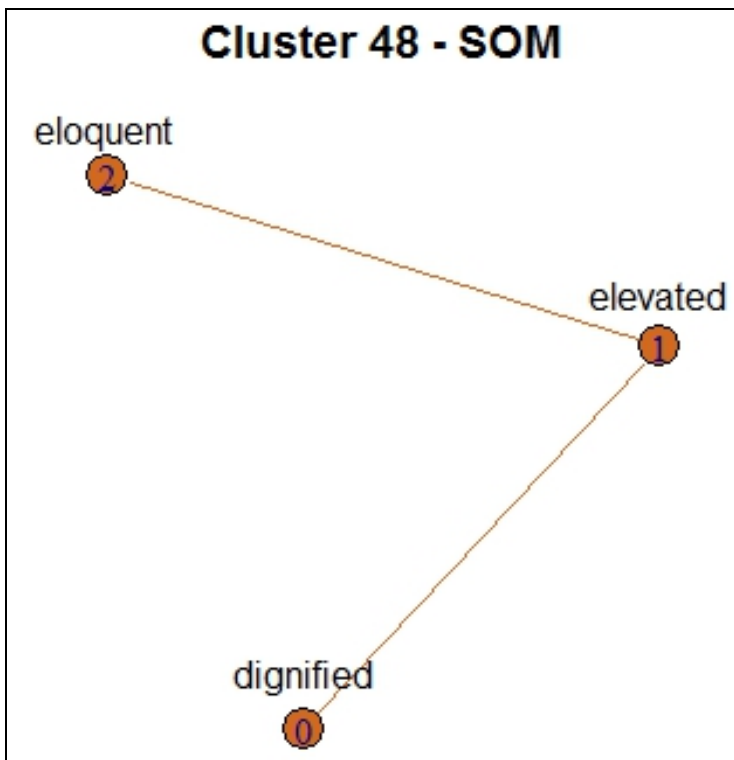


Figura 61: *Cluster 48* – Metodologia SOM

Nota: idem da nota da figura 1.

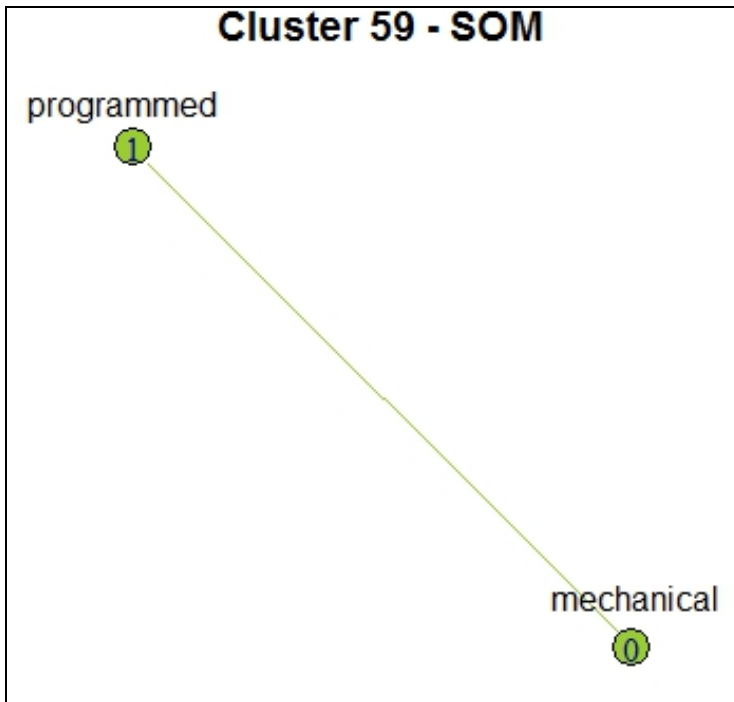


Figura 62: *Cluster 59* – Metodologia SOM

Nota: idem da nota da figura 1.

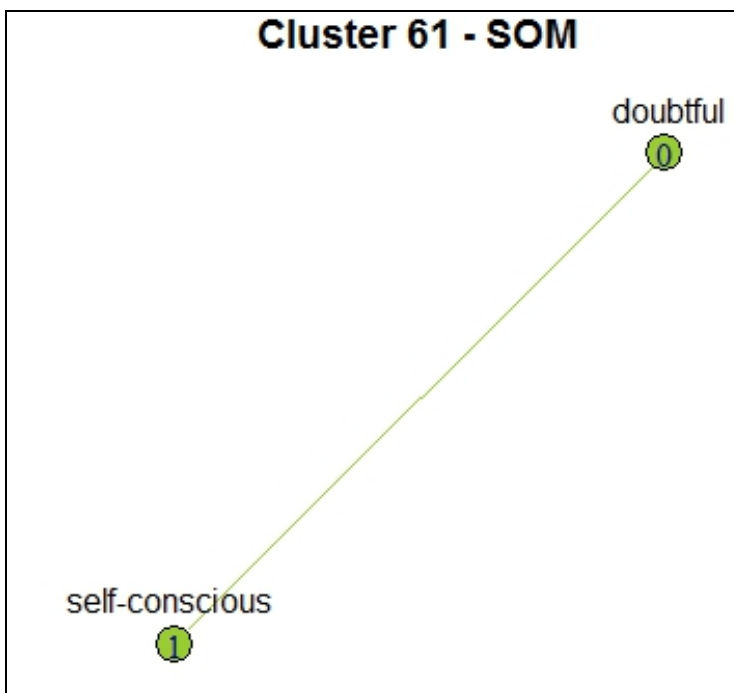


Figura 63: *Cluster 61* – Metodologia SOM

Nota: idem da nota da figura 1.

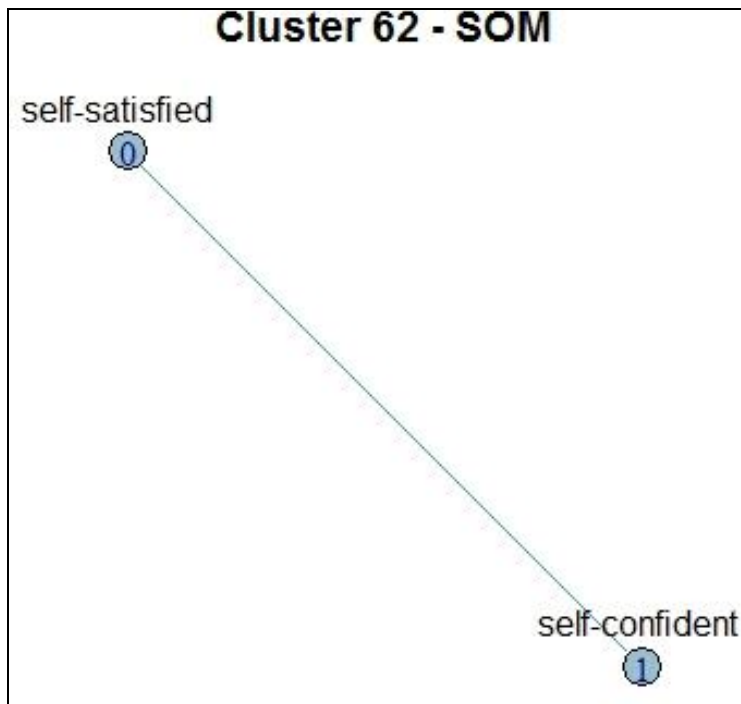


Figura 64: *Cluster 62* – Metodologia SOM

Nota: idem da nota da figura 1.

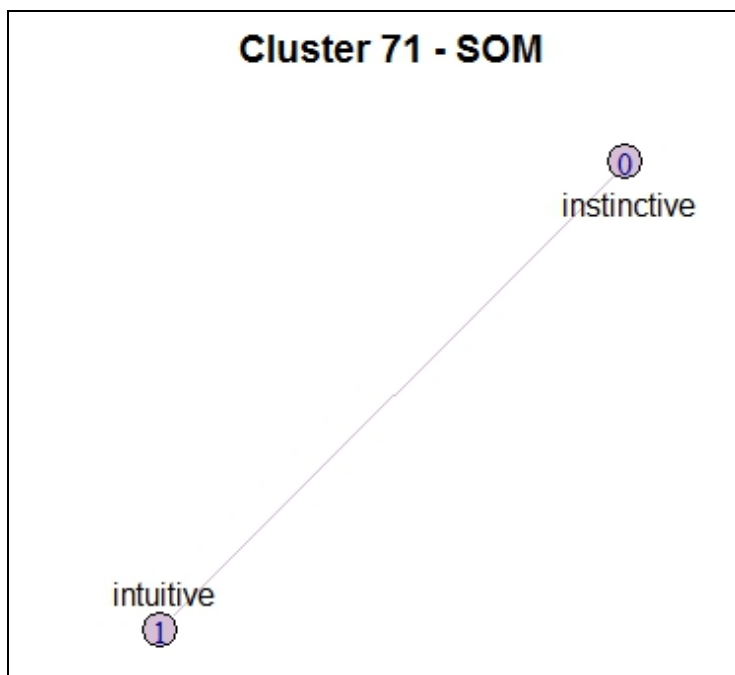


Figura 65: *Cluster 71* – Metodologia SOM

Nota: idem da nota da figura 1.

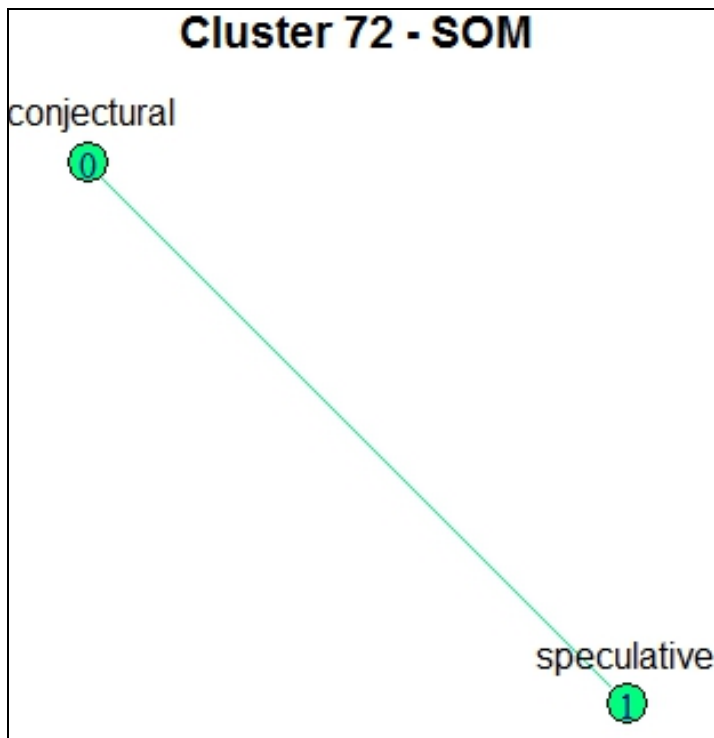


Figura 66: *Cluster 72* – Metodologia SOM

Nota: idem da nota da figura 1.

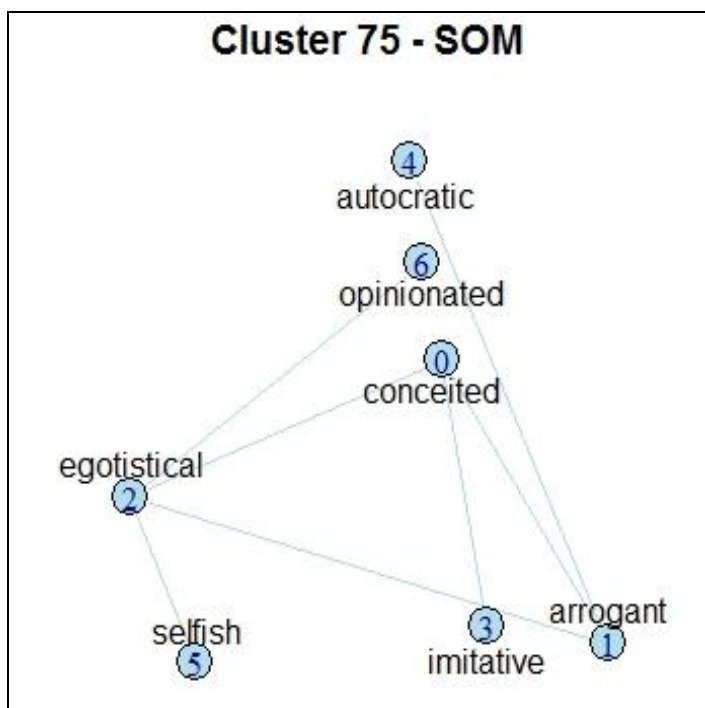


Figura 67: *Cluster 75* – Metodologia SOM

Nota: idem da nota da figura 1.

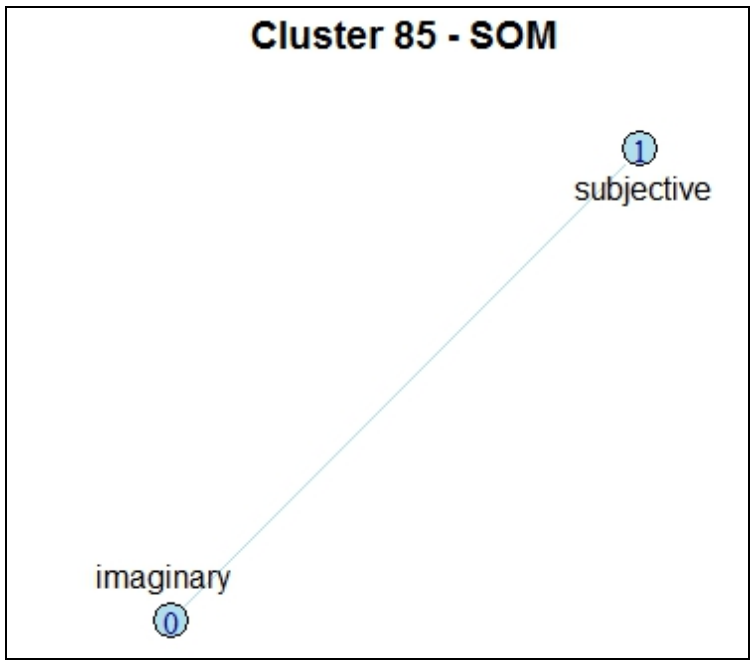


Figura 68: *Cluster 85* – Metodologia SOM

Nota: idem da nota da figura 1.

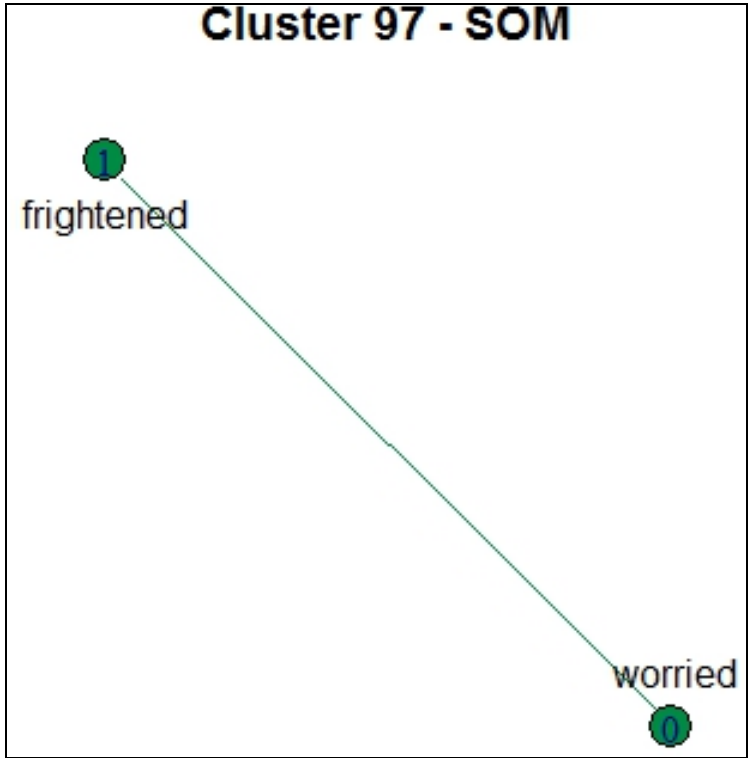


Figura 69: *Cluster 97* – Metodologia SOM

Nota: idem da nota da figura 1.

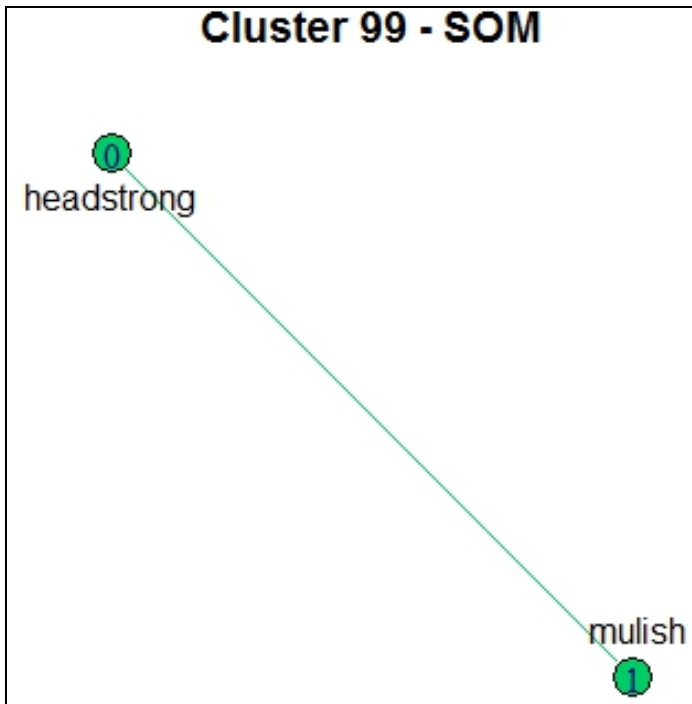


Figura 70: Cluster 99 – Metodologia SOM

Nota: idem da nota da figura 1.

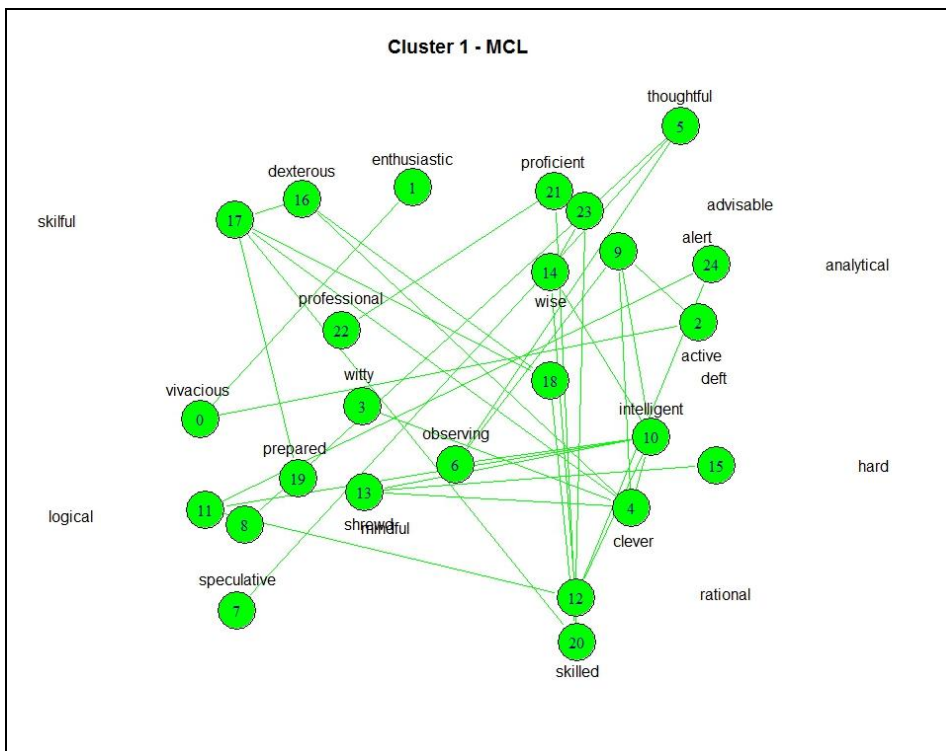


Figura 71: Cluster 1 – Metodologia MCL

Nota: idem da nota da figura 1.

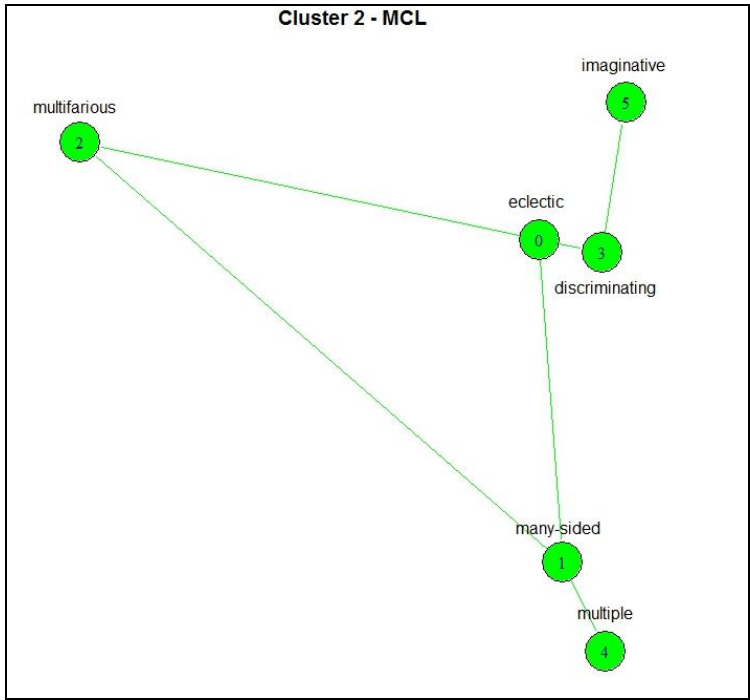


Figura 72: Cluster 2 – Metodologia MCL

Nota: idem da nota da figura 1.

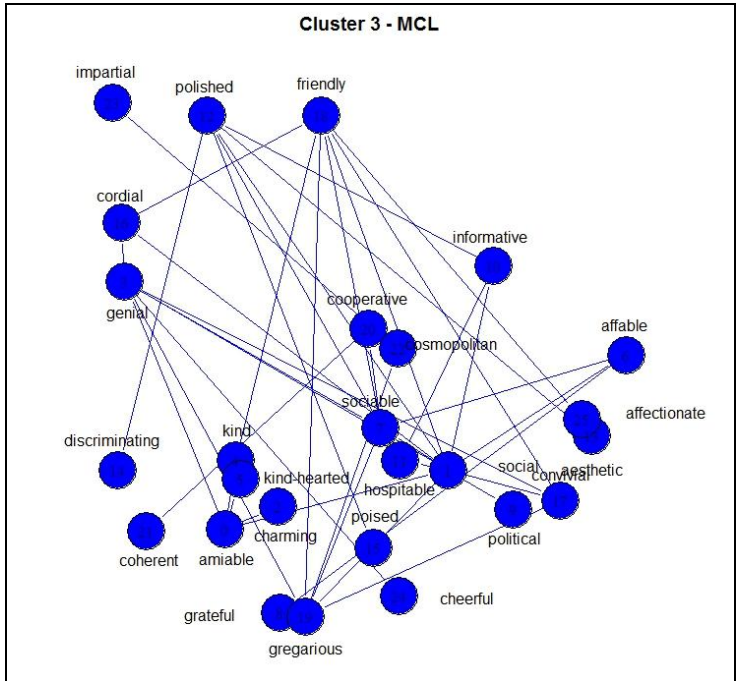


Figura 73: Cluster 3 – Metodologia MCL

Nota: idem da nota da figura 1.

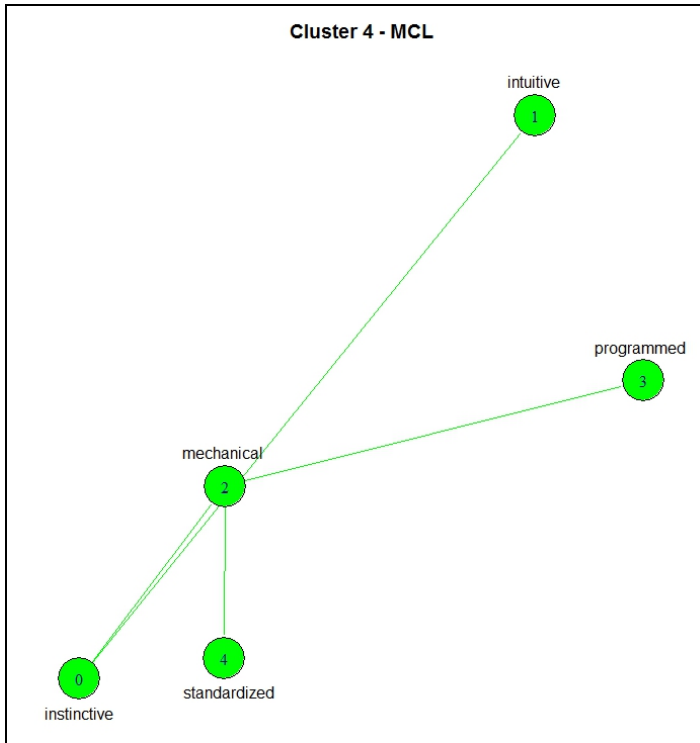


Figura 74: Cluster 4 – Metodologia MCL

Nota: idem da nota da figura 1.

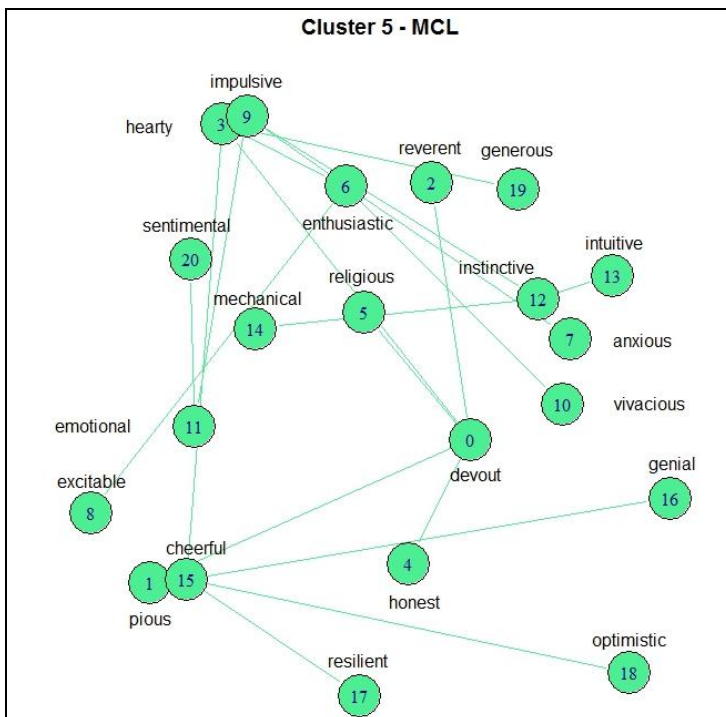


Figura 75: Cluster 5 – Metodologia MCL

Nota: idem da nota da figura 1.

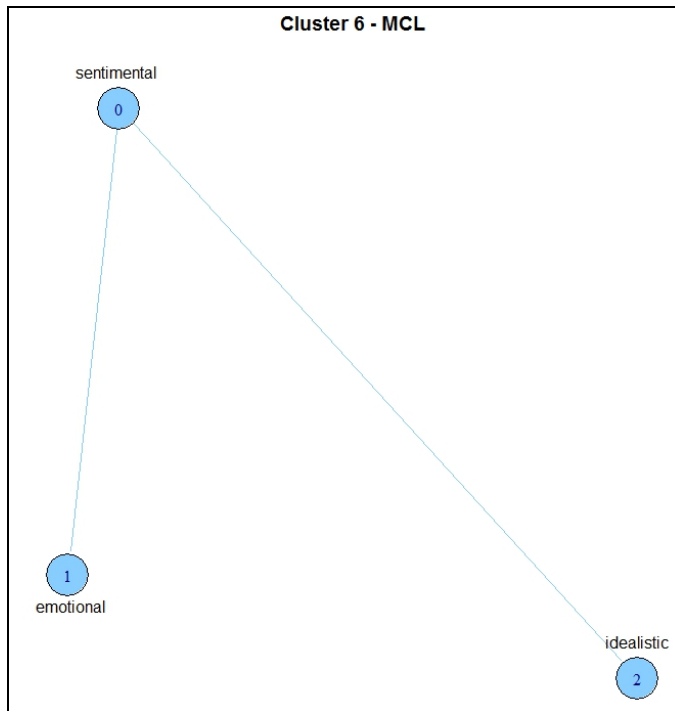


Figura 76: Cluster 6 – Metodologia MCL

Nota: idem da nota da figura 1.

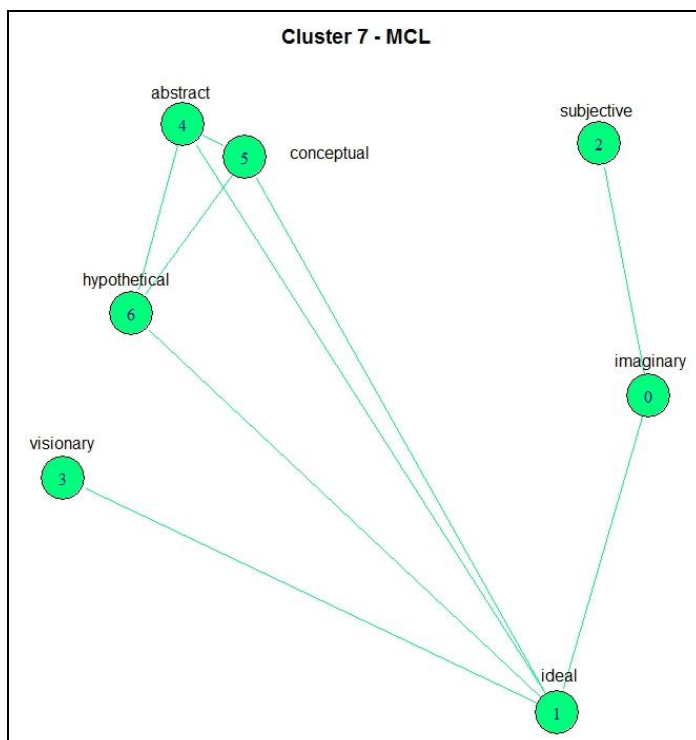


Figura 77: Cluster 7 – Metodologia MCL

Nota: idem da nota da figura 1.



Figura 78: Cluster 8 – Metodologia MCL

Nota: idem da nota da figura 1.

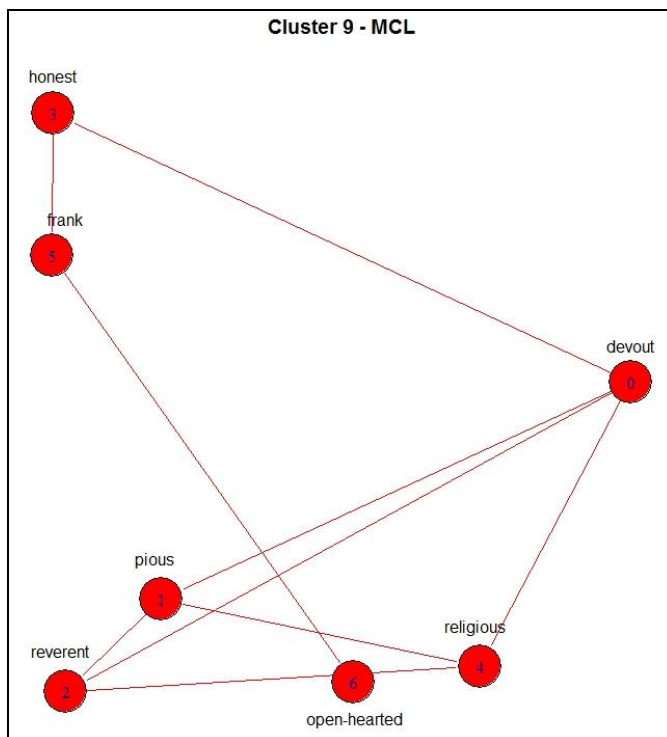


Figura 79: Cluster 9 – Metodologia MCL

Nota: idem da nota da figura 1.

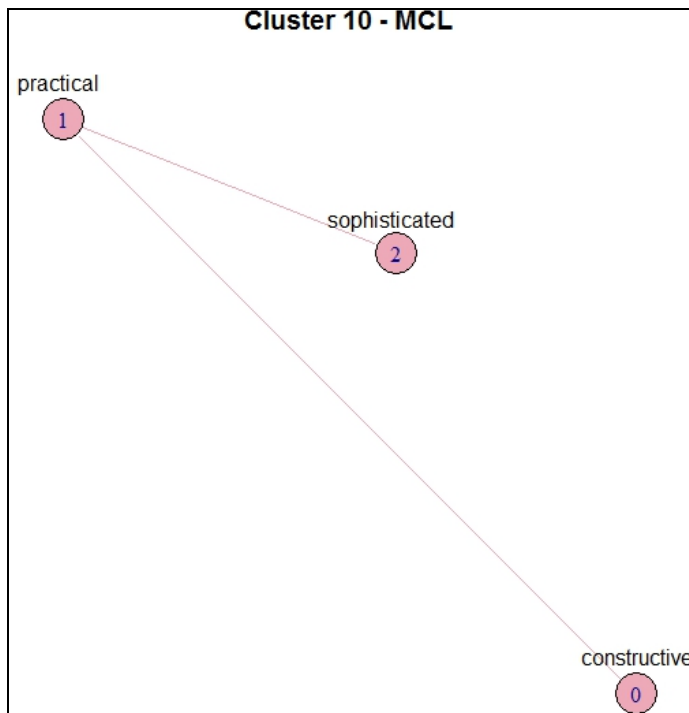


Figura 80: *Cluster 10* – Metodologia MCL

Nota: idem da nota da figura 1.

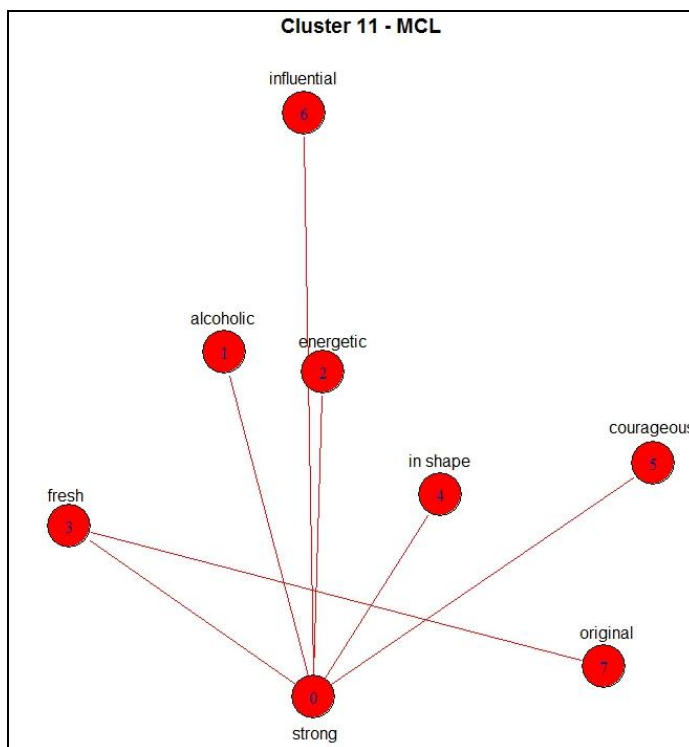


Figura 81: *Cluster 11* – Metodologia MCL

Nota: idem da nota da figura 1.

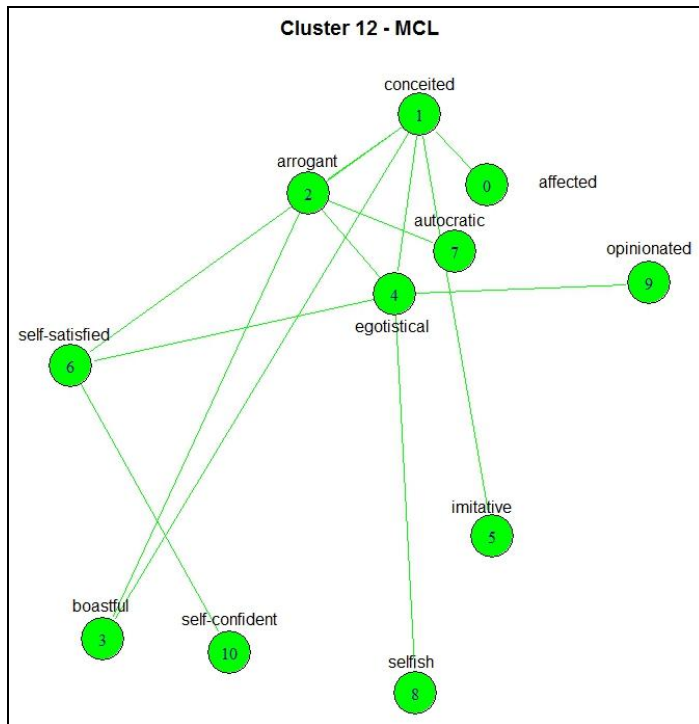


Figura 82: *Cluster 12* – Metodologia MCL

Nota: idem da nota da figura 1.

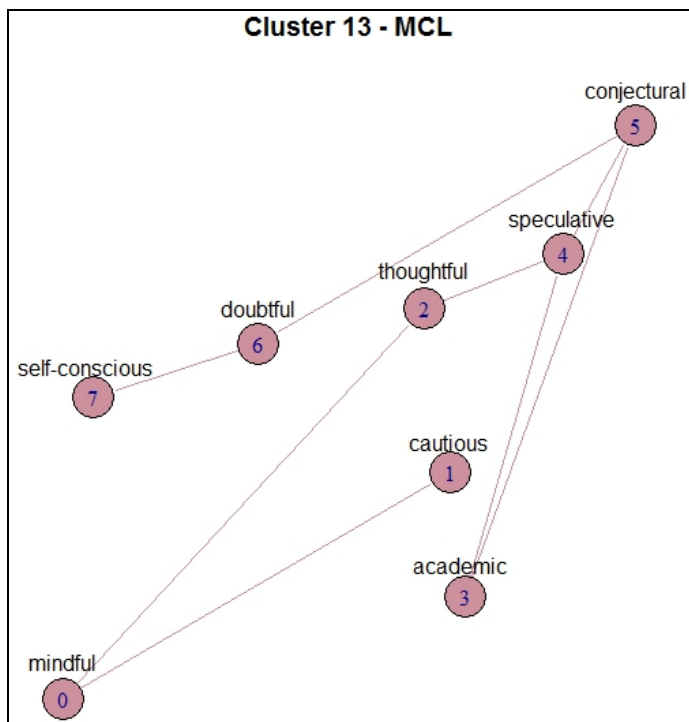


Figura 83: *Cluster 13* – Metodologia MCL

Nota: idem da nota da figura 1.

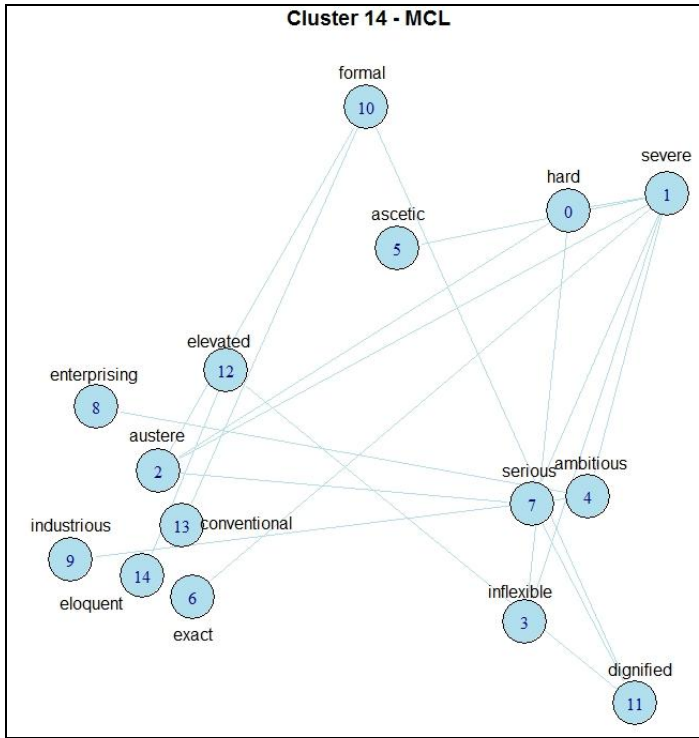


Figura 84: Cluster 14 – Metodologia MCL

Nota: idem da nota da figura 1.

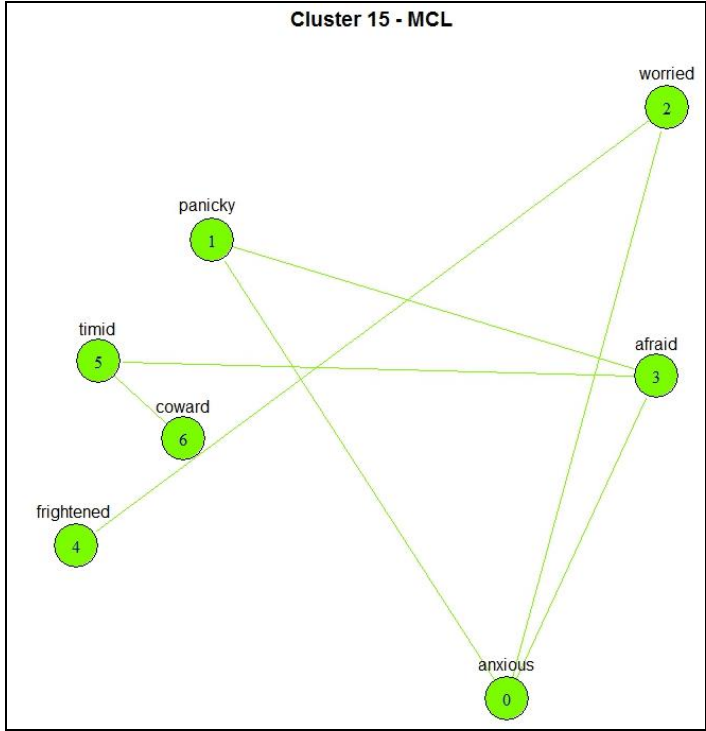


Figura 85: Cluster 15 – Metodologia MCL

Nota: idem da nota da figura 1.

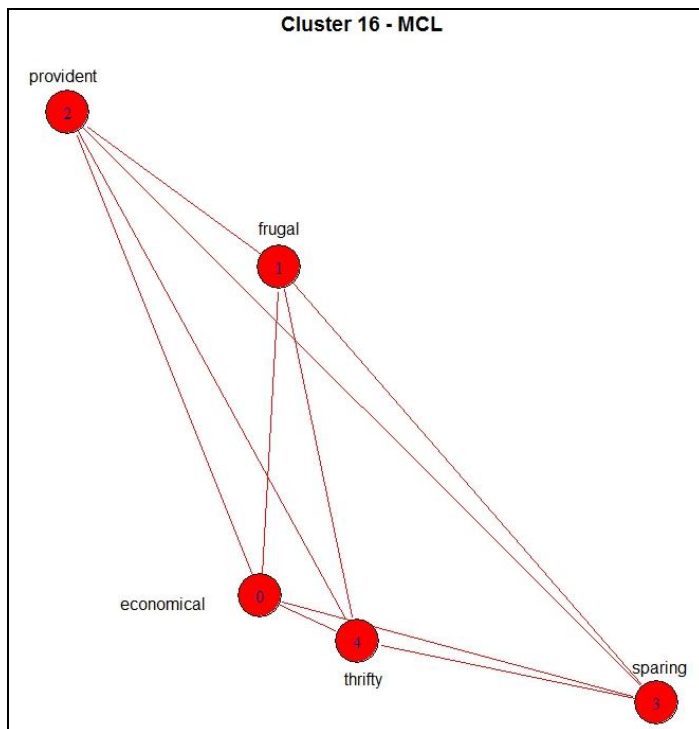


Figura 86: Cluster 16 – Metodologia MCL

Nota: idem da nota da figura 1.

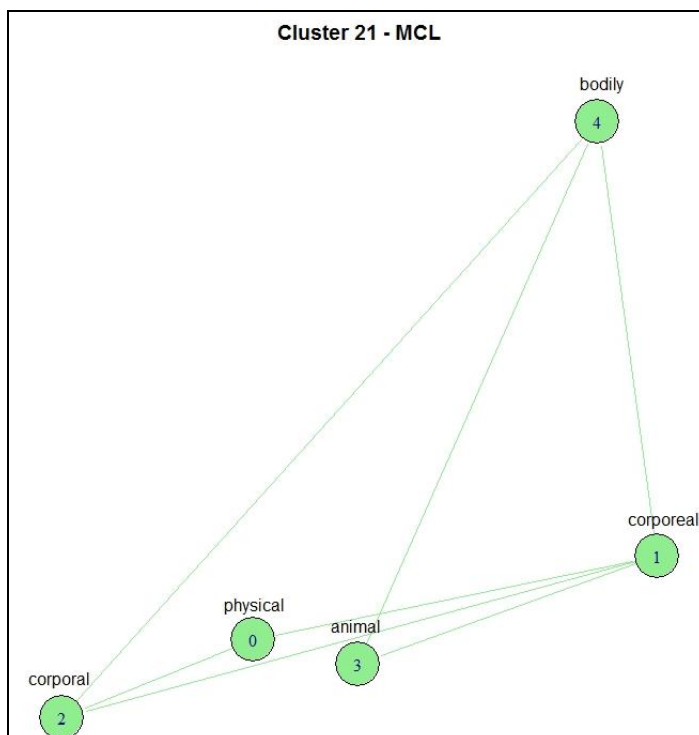


Figura 87: Cluster 21 – Metodologia MCL

Nota: idem da nota da figura 1.

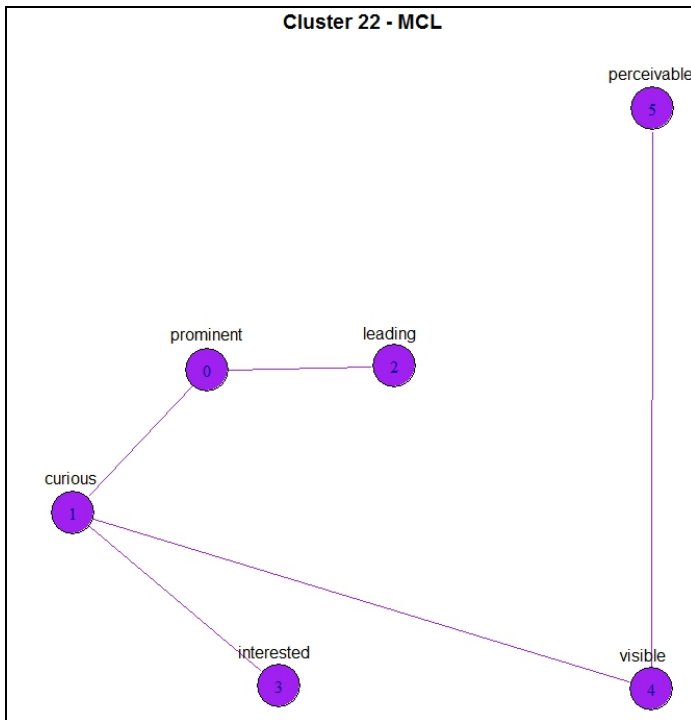


Figura 88: *Cluster 22* – Metodologia MCL

Nota: idem da nota da figura 1.

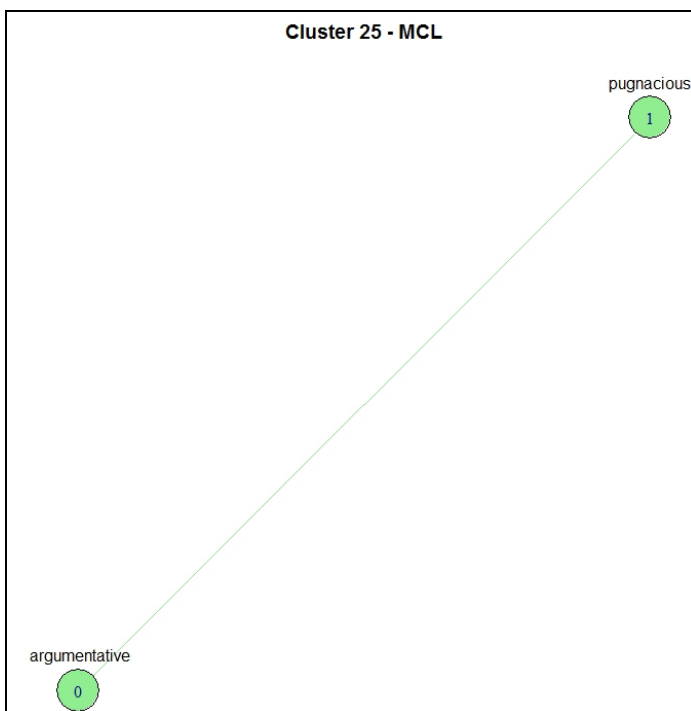


Figura 89: *Cluster 25* – Metodologia MCL

Nota: idem da nota da figura 1.

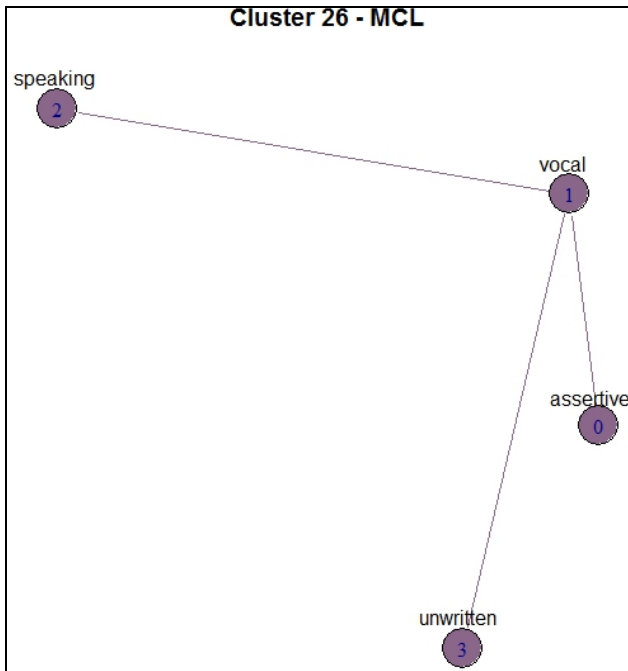


Figura 90: Cluster 26 – Metodologia MCL

Nota: idem da nota da figura 1.

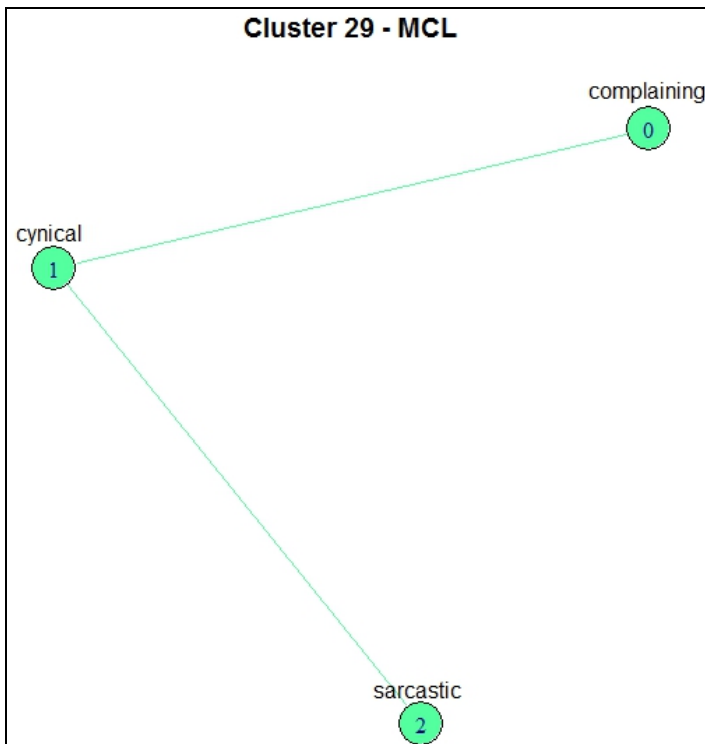


Figura 91: Cluster 29 – Metodologia MCL

Nota: idem da nota da figura 1.

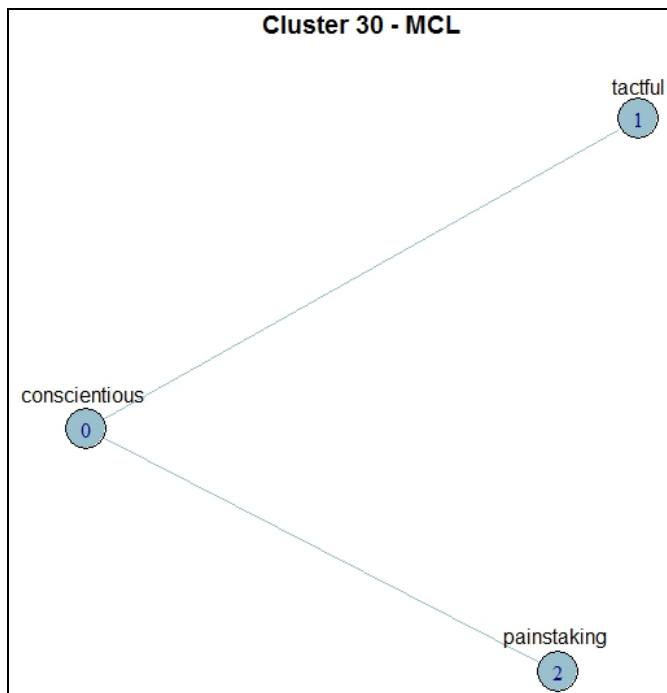


Figura 92: *Cluster 30* – Metodologia MCL

Nota: idem da nota da figura 1.

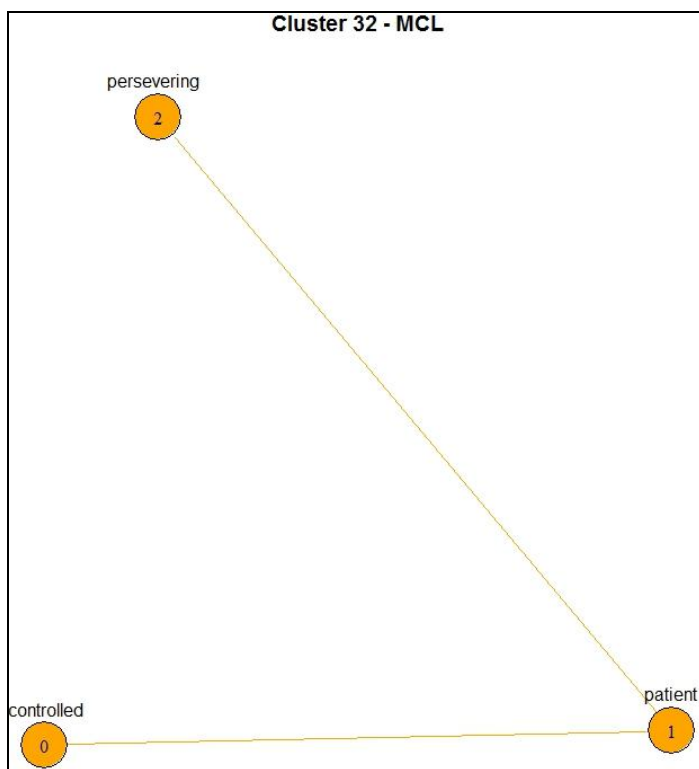


Figura 93: *Cluster 32* – Metodologia MCL

Nota: idem da nota da figura 1.

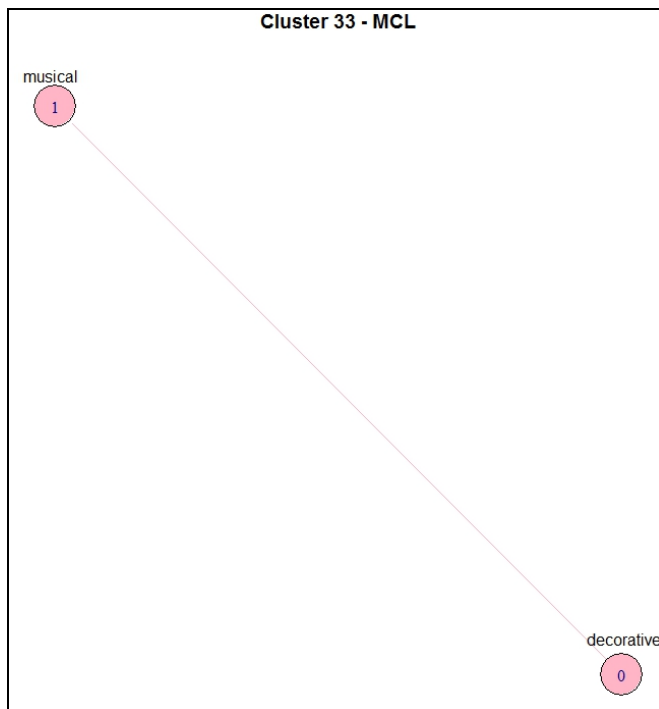


Figura 94: *Cluster 33* – Metodologia MCL

Nota: idem da nota da figura 1.

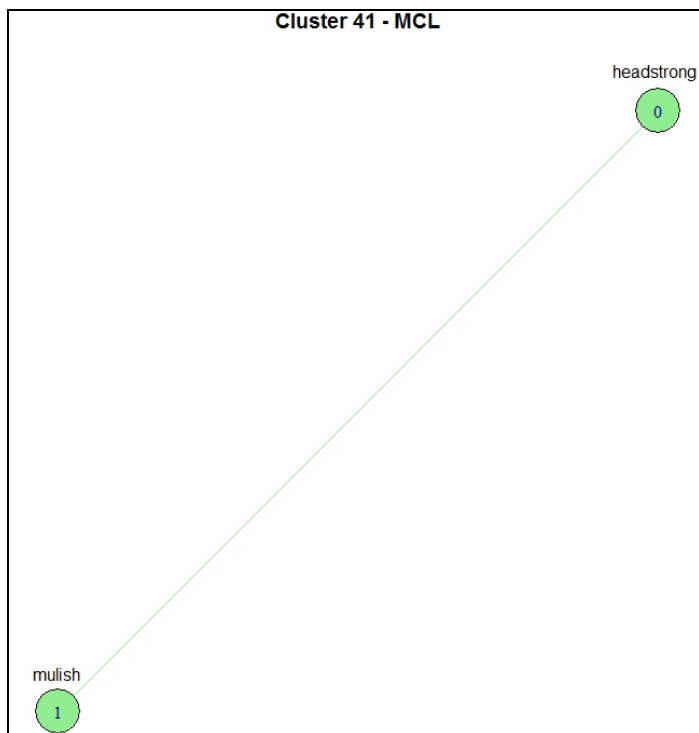


Figura 95: *Cluster 41* – Metodologia MCL

Nota: idem da nota da figura 1.

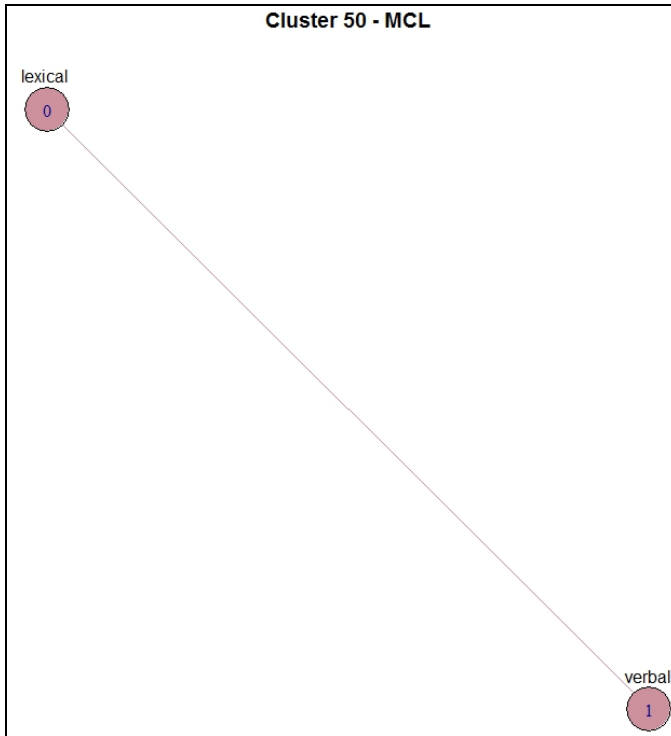


Figura 96: *Cluster 50* – Metodologia MCL

Nota: idem da nota da figura 1.

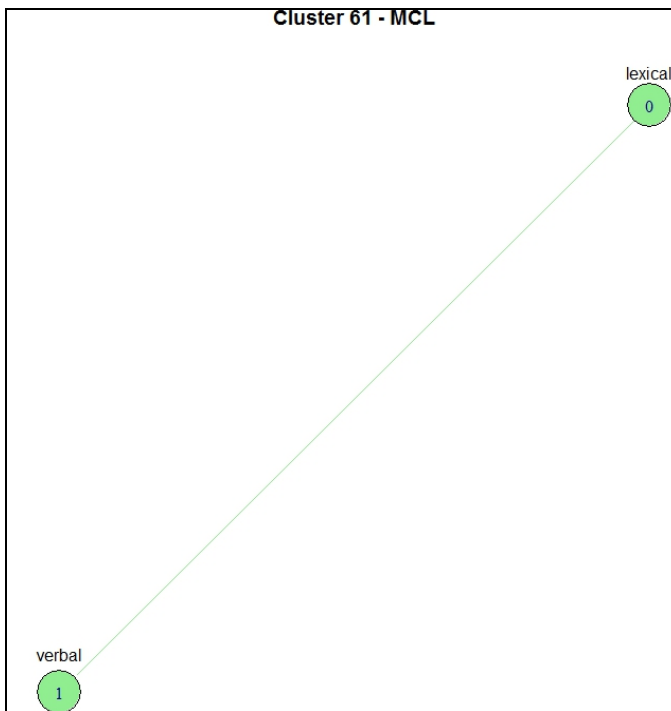


Figura 97: *Cluster 61* – Metodologia MCL

Nota: idem da nota da figura 1.

ANEXO

Tabela 41: Carta do linguista da empresa True Term Thesaurus Database, recebida em 25-11-2013 confirmando que maioria das ligações diretas entre palavras neste thesaurus correspondem a sinônimos.

attached the answer from our linguist.

Best Regards

Wolfgang

*Wolfgang Fritscher
TT-Software/Databases
www.language-databases.com
www.true-term-mobile.com*

hello, again.

i can not see any problems with the product you got from true-term, here: thesaurus, english.

the purpose of a thesaurus is to offer the user a listing of similar or related words(<http://en.wikipedia.org/wiki/Synonym>, on november/25/2013).

the synonyms(etc) are given after the sign ":";

the (primary) function of the word(s) before the sign ":" are not "to be or not to be a synonym"-nevertheless "these are often, but not always, synonyms".

this belongs to nearly all words except ca 120 words-please see my file: 120 words do not have a follower after the sign ":"!

for this words i made a proofreading etc.

how to handle these words:

first: throw it away

or

second: see sheet 2

or

third: expand the thesaurus.

get contact for further questions.