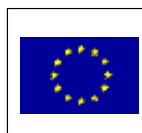


Pedro Miguel Lopes Bastos

# Inferência de propriedades químicas do algodão através de técnicas de Data Mining

Dissertação submetida à Universidade do Minho para a obtenção do grau de Mestre em Sistemas de Informação



Universidade do Minho

2003



Aos meus pais, à minha avó e à  
Lillian



## Agradecimentos

Desejo agradecer a todos os que, de alguma forma, contribuíram para a realização deste trabalho, em particular.

Aos Profs. Luís Amaral e Rosa Vasconcelos, pela supervisão desta tese, bem como pelos conselhos e sugestões dadas ao longo da realização deste trabalho.

À Prof. Maribel Santos, pelos conselhos e sugestões dadas no domínio do Data Mining e da ferramenta utilizada.

À Dr<sup>a</sup> Sivia Dias, por todo o apoio e ajuda dada no decorrer da elaboração deste trabalho.

A todos os restantes elementos pertencentes ao projecto de investigação, no qual este trabalho está inserido, em particular à Prof. Teresa Amorin.

Ao Eng<sup>o</sup> Luís Pires pelo apoio e conselhos dados no decorrer deste trabalho.

Aos meus restantes colegas do CIESA, pelo incentivo fornecido.

Aos meus pais e avó, por sempre me incentivarem e apoiarem durante toda a minha formação, quer académica como profissional.

À memória do meu avô, pois embora o seu apoio no início desta caminhada, não pode assistir à sua conclusão.

À minha família, pelo incentivo e apoio dado ao longo dos anos.

E finalmente, queria agradecer à Lillian, a paciência, o apoio e o incentivo dado em todas as horas.



Esta dissertação está inserida no âmbito do projecto  
POCTI/1999/CTM/32993, financiado pela  
FCT (Fundação para a Ciência e Tecnologia).



## Resumo

Este trabalho descreve a forma como a ferramenta de prospecção de dados (*data mining*) *Clementine* pode ser utilizada na extração de conhecimento de dados relativos a propriedades físicas e químicas da fibra de algodão. Os resultados atingidos demonstram a forma como as técnicas de prospecção de dados podem ser usadas para estabelecer, de forma eficiente, relações existentes entre as propriedades das fibras.

O desenvolvimento tecnológico tornou possível a medição das diferentes propriedades físicas das fibras de algodão tipo comprimento, micronaire, grau de uniformidade, alongamento, resistência, cor e grau de impureza. Isto é conseguido através da utilização de instrumentos HVI<sup>1</sup>, proporcionando resultados rápidos e fiáveis. No entanto, no que diz respeito ao estudo das propriedades químicas os resultados são obtidos usando métodos laboratoriais mais demorados e dispendiosos, assim por vezes são completamente ignorados pelos diferentes agentes englobados no processo de transformação da matéria prima em produto final. Assim sendo, o estudo de todas as possíveis relações existentes entre as diferentes propriedades físicas e químicas são descartadas.

Este conhecimento é bastante importante, pois as propriedades químicas afectam em muito o processo de transformação das referidas fibras. Desta forma, através da utilização do *Clementine*, é possível obter relações entre os diferentes tipos de características da fibra apoiadas na geração de regras utilizando algoritmos de inteligência artificial.

Neste estudo, são usadas várias técnicas de prospecção de dados existentes no *Clementine*. A prospecção de dados consiste num dos passos do processo de ECBD<sup>2</sup>, processo que tem como principal função a descoberta de conhecimento entre conjuntos de dados. A ferramenta inclui técnicas avançadas de modelação baseadas em inteligência artificial, extraíndo dos dados, possíveis relações complexas existentes, bem como regras de associação entre eles. Isto ajuda a automatizar processos tipo predição, estimativa e classificação, que podem ser usadas como forma de proporcionar decisões de suporte especializado.

---

<sup>1</sup> High Volume Instrumentes Systems

<sup>2</sup> Extração de Conhecimento de Bases de Dados



## **Abstract**

This work describes how the data mining tool named Clementine, can be used in knowleged extraction of phisical and chemical properties in cotton fibre. The resultes achieved demonstrate how the data mining tools can be used to establish, in a efficient way, all the existent relations between the fibre properties.

The technological development enabled the measurement of cotton staple fibres properties like length, micronaire, uniformity, strength, elongation, color and trash contents. This is obtained through the use of HVI instruments, providing rapid and reliable results. However, with regard to the study of the chemical properties, the results are obtained by using more time consuming and much more expensive laboratory methods, although, sometimes they are completely ignored by all the agents involved in the transformation process. This means that the studies about all existent relationships between the fisical and chemical properties are discarded.

This knowledge is very important, because chemical properties affect the process of fiber transformation. In this way, by using Clementine it is possible to obtain relations between diferent types of cotton fibres, supported by the creation of rules using artificial intelligence.

In this study, several data mining techniques available in the Clementine data mining system are used. Data mining consists of a step in the knowledge discovery process (KDD), a process that aims to discover associations between data sets. The tool includes advanced modelling techniques, based in artificial intelligence, extracting of the existing data, complex relationships, as well as possible association rules between them. This helps to automatize processes such as prediction, estimation and classification, that can be used to provide expert decision support.



**Palavras-chave:** Extração de Conhecimento em Bases de Dados, Prospecção de Dados (*Data Mining*), Propriedades do Algodão.

**Key Words:** Knowledge Discovery in Databases, Data Mining, Cotton Properties.



# Índice

INTRODUÇÃO .....	1
1.1 - Contexto .....	1
1.2 - Objectivos .....	2
1.3 - Motivação .....	2
1.4 - Estrutura da Dissertação .....	3
A EXTRACÇÃO DE CONHECIMENTO NAS BASES DE DADOS.....	5
2.1 - Visão geral .....	5
2.2 - Antecedentes .....	6
2.3 - Em que consiste a prospecção de dados ( <i>data mining</i> )?.....	7
2.4 - O ECBD, prospecção de dados e a relação com outros campos.....	7
2.5 - Prospecção de dados, definição e conceitos gerais.....	8
2.6 - Processo de ECBD.....	9
2.6.1 - Compreensão do Domínio de Aplicação .....	12
2.6.2 - Selecção do Dados .....	13
2.6.3 - Pré-processamento dos dados .....	13
2.6.4 - Data Mining ou Prospecção de Dados .....	16
2.6.4.1 - <i>Funções da prospecção de dados</i> .....	17
2.6.4.2 - <i>Algoritmos usados na prospecção de dados</i> .....	18
2.6.4.3 - <i>Extracção de Padrões</i> .....	21
2.6.5 - Interpretação e Avaliação do Conhecimento .....	22
2.6.6- Consolidação do Conhecimento .....	22
2.7 - Considerações Finais .....	23
A FIBRA DE ALGODÃO .....	24
3.1 - Nota Histórica .....	24
3.2 - Morfologia da Planta.....	25
3.3 - Fisiologia da Planta e os Factores que influenciam o seu desenvolvimento .....	26
3.4 - Classificação e Padronização.....	28
3.5 - Estrutura morfológica da fibra .....	31
3.5.1 - Corte Longitudinal .....	33
3.5.2 - Corte Transversal.....	33
3.6 - Propriedades Físicas.....	34
3.6.1 – Comprimento.....	34
3.6.2 - Propriedades Dinamométricas .....	38
3.6.2.1 - <i>Resistência e Tenacidade</i> .....	39
3.6.2.2 - <i>Alongamento</i> .....	41



3.6.3 - Finura e Maturidade .....	41
3.6.4 - Cor .....	44
3.7 - Características Químicas .....	45
3.7.1 - Celulose .....	46
3.7.2 - Pectinas .....	47
3.7.3 - Substâncias proteicas .....	48
3.7.4 - Ceras .....	48
3.7.5 - Cinzas .....	49
3.7.6 - Açúcares .....	49
3.8 - Considerações finais .....	50
DESCRIÇÃO DA FERRAMENTA .....	51
4.1 - Introdução .....	51
4.2 - Técnicas de modelação .....	52
4.2.1 - Redes neuronais .....	52
4.2.2 - Indução de regras .....	53
4.2.3 - Redes de Kohonen .....	54
4.2.4 - Regras de associação .....	55
4.3 - Interface do Clementine .....	56
4.4 - Como utilizar o Clementine .....	57
4.4.1 - Descrição dos ícones .....	59
4.4.2 - Criação da <i>Stream</i> .....	63
4.4.3 - Super-nós .....	63
4.5 - Considerações finais .....	65
TRABALHO PRÁTICO .....	66
5.1 - Selecção e Preparação das Bases de Dados .....	66
5.2 - Criação de escalões das propriedades químicas em estudo .....	69
5.3 - Relação entre as propriedades químicas e a origem da fibra .....	71
5.4 - Importância das propriedades físicas .....	73
5.5 - Relações entre as propriedades químicas e físicas .....	74
5.5.1 - %Cinzas (Percentagem de cinzas) .....	74
5.5.1.1 - Resultados obtidos .....	76
5.5.2 - Ca (Cálcio) .....	77
5.5.2.1 - Resultados obtidos .....	78
5.5.3 - K (Potássio) .....	79
5.5.3.1 - Resultados obtidos .....	80
5.5.4 - ME (Materiais Extraídos) .....	80
5.5.4.1 - Resultados obtidos .....	81
5.5.5 - Mg (Magnésio) .....	82
5.5.5.1 - Resultados obtidos .....	82
5.5.6 - P (Fósforo) .....	83
5.5.6.1 - Resultados obtidos .....	84



5.6 - Discussão de Resultados .....	85
5.7 – Considerações finais.....	89
<b>CONCLUSÕES E TRABALHO FUTURO .....</b>	<b>90</b>
6.1 – Conclusões .....	90
6.2 – Trabalho Futuro.....	94



## Índice de Ilustrações

<b>FIGURA 1</b> – VISÃO GENERALISTA DO PROCESSO DE ECBD (ABONYI ET AL., 2000).....	10
<b>FIGURA 2</b> – ETAPAS CONSTITUINTES DO PRÉ-PROCESAMENTO DOS DADOS (HAN ET AL., 2001).....	14
<b>TABELA 1</b> – CARACTERÍSTICAS GERAIS DOS SISTEMAS DE APRENDIZAGEM MÁQUINA (KOHAVIC ET AL., 1994). ....	18
<b>TABELA 2</b> – CLASSIFICAÇÃO DAS FIBRAS QUANTO AO COMPRIMENTO (VASCONCELOS, 1993).....	36
<b>TABELA 3</b> – CLASSIFICAÇÃO DAS FIBRAS QUANTO AO ÍNDICE DE UNIFORMIDADE (VASCONCELOS, 1993).....	36
<b>TABELA 4</b> – CLASSIFICAÇÃO DO ALGODÃO QUANTO À RESISTÊNCIA (KLEIN, 1980). ....	39
<b>TABELA 5</b> – CLASSIFICAÇÃO DAS FIBRAS QUANTO À TENACIDADE (VASCONCELOS, 1993). ....	40
<b>TABELA 6</b> – CLASSIFICAÇÃO DO ALGODÃO QUANTO AO ALONGAMENTO (PARKER, 1982). ....	41
<b>EQUAÇÃO 1</b> – DEFINIÇÃO DE MATURIDADE (VASCONCELOS, 1993).....	43
<b>TABELA 7</b> – CLASSIFICAÇÃO DO ALGODÃO QUANTO À MATURIDADE (VASCONCELOS, 1993). ....	43
<b>TABELA 8</b> – CLASSIFICAÇÃO DO ALGODÃO QUANTO AO ÍNDICE MICRONAIRE (FERREIRA, 1960). ....	44
<b>TABELA 9</b> – CARACTERÍSTICAS DE ALGODÕES DE DIFERENTES ORIGENS, EM MATÉRIA SECA (VASCONCELOS, 1993). ....	46
<b>TABELA 10</b> – COMPOSIÇÃO DA CAMADA PRIMÁRIA E DA FIBRA, EM PORCENTAGEM (LUCAS, 2000). ....	47
<b>FIGURA 3</b> – REDE NEURONAL .....	52
<b>FIGURA 4</b> – ÁRVORE DE DECISÃO .....	53
<b>FIGURA 5</b> – REDES DE <i>KOHONEN</i> .....	54
<b>FIGURA 6</b> – JANELA PRINCIPAL DO CLEMENTINE COM TODOS OS SEUS COMPONENTES.....	56
<b>FIGURA 7</b> – EXEMPLO DE UMA <i>STREAM</i> .....	58
<b>FIGURA 8</b> – EXCERTO DA TABELA DE DADOS REFERENTES ÀS PROPRIEDADES FÍSICAS. ....	67
<b>FIGURA 9</b> – <i>STREAM SEPARACAO.STR</i> .....	68
<b>FIGURA 10</b> – <i>STREAM CRIA_ESC_QUIM.STR</i> .....	69
<b>TABELA 11</b> – ESCALÕES DE % CINZAS GERADOS.....	70
<b>TABELA 12</b> – ESCALÕES DE % ME GERADOS .....	70
<b>TABELA 13</b> – ESCALÕES DE K GERADOS .....	70
<b>TABELA 14</b> – ESCALÕES DE MG GERADOS.....	70
<b>TABELA 15</b> – ESCALÕES DE P GERADOS. ....	70
<b>TABELA 16</b> – ESCALÕES DE CA GERADOS. ....	70
<b>FIGURA 11</b> – <i>STREAM RELACAO.STR</i> (UTILIZANDO O MODELO C5.0).....	71
<b>FIGURA 12</b> – <i>STREAM RELACAO.STR</i> (UTILIZANDO O MODELO GRI).....	72
<b>TABELA 17</b> – ESCALÕES QUÍMICOS POR ORIGEM. ....	73
<b>FIGURA 13</b> – <i>STREAM IMPORTANCIA.STR</i> .....	74
<b>FIGURA 14</b> – <i>STREAM ESC_TOTAL.STR</i> .....	75
<b>TABELA 18</b> – ESCALÕES DE +B GERADOS. ....	75
<b>TABELA 19</b> – ESCALÕES DE MICRONAIRE GERADOS. ....	75
<b>TABELA 20</b> – ESCALÕES DE RD GERADOS.....	76
<b>FIGURA 15</b> – <i>STREAM ANALISE_RELACOES.STR</i> .....	76



<b>TABELA 21</b> – MATRIZ ESCALÕES DE % CINZAS VS. PROPRIEDADES FÍSICAS RELEVANTES. ....	77
<b>TABELA 22</b> – ESCALÕES DE UI GERADOS. ....	78
<b>TABELA 23</b> – ESCALÕES DE ST GERADOS. ....	78
<b>TABELA 24</b> – MATRIZ ESCALÕES DE CA VS. PROPRIEDADES FÍSICAS RELEVANTES. ....	79
<b>TABELA 25</b> – ESCALÕES DE SL50 GERADOS. ....	79
<b>TABELA 26</b> – MATRIZ ESCALÕES DE K VS. PROPRIEDADES FÍSICAS RELEVANTES. ....	80
<b>TABELA 27</b> – ESCALÕES DE PSI GERADOS. ....	81
<b>TABELA 28</b> – MATRIZ ESCALÕES DE ME VS. PROPRIEDADES FÍSICAS RELEVANTES. ....	82
<b>TABELA 29</b> – ESCALÕES DE SL2.5 GERADOS. ....	82
<b>TABELA 30</b> – MATRIZ ESCALÕES DE MG VS. PROPRIEDADES FÍSICAS RELEVANTES. ....	83
<b>TABELA 31</b> – ESCALÕES DE EL GERADOS. ....	84
<b>TABELA 32</b> – MATRIZ ESCALÕES DE P VS. PROPRIEDADES FÍSICAS RELEVANTES. ....	84
<b>TABELA 33</b> – MATRIZ +B (GRAU DE AMARELO). ....	85
<b>TABELA 34</b> – MATRIZ RD (REFLECTÂNCIA). ....	85
<b>TABELA 35</b> – MATRIZ MICRONAIRE. ....	86
<b>TABELA 36</b> – MATRIZ PSI (RESISTÊNCIA). ....	86
<b>TABELA 37</b> – MATRIZ ST (RESISTÊNCIA). ....	87
<b>TABELA 38</b> – MATRIZ SL2.5 (COMPRIMENTO). ....	87
<b>TABELA 39</b> – MATRIZ UI (ÍNDICE DE UNIFORMIDADE). ....	87
<b>TABELA 40</b> – MATRIZ SL50 (COMPRIMENTO). ....	88
<b>TABELA 41</b> – MATRIZ EL (ALONGAMENTO). ....	88
<b>TABELA 42</b> – INFLUÊNCIA ENTRE AS PROPRIEDADES FÍSICAS E QUÍMICAS. ....	89



# Cap. 1

## INTRODUÇÃO

---

### 1.1 - CONTEXTO

A capacidade de armazenamento em bases de dados, assim como a sua utilização, encontra-se em crescimento numa proporção simétrica aos avanços em novas tecnologias de informação e comunicação. A actividade de extracção de informações relevantes está, por conseguinte, a tornar-se bastante complexa. Este processo de “extracção” é denominado por *Knowledge Discovery in Databases* – KDD (Extracção de Conhecimento de Bases de Dados).

O evoluir tecnológico, quer na área do *hardware*, quer na área do *software*, possibilitou a criação de aplicações capazes de processar enormes quantidades de dados. Cada vez mais as organizações têm investido em equipamentos e aplicações capazes de armazenar, gerir e entender os seus dados. Isto porque, hoje em dia, as organizações não olham para as bases de dados como simples repositório de informação, mas sim como um importante património organizacional (Sen et al., 1998).

A quantidade de dados gerados por um qualquer sistema organizacional, pode por vezes superar a capacidade humana de interpretar e compreender tanta informação. Assim, torna-se necessário a utilização de técnicas e ferramentas capazes de analisar de forma automática os dados produzidos e armazenados, fornecendo o conhecimento necessário de forma a auxiliar os processos de tomada de decisão (Oliveira, 2000).

A área da computação conhecida por Extracção de Conhecimento de Bases de Dados (ECBD) ou *Knowledge Discovery in Databases* (KDD), surge como forma de auxílio na análise de grandes quantidades de dados. As pesquisas nesta área têm como objectivo o estudo da aplicação de novas metodologias, técnicas e ferramentas capazes da extracção de conhecimentos implícito em grandes quantidades de dados.

O processo de ECBD utiliza conceitos de bases de dados, métodos estatísticos, ferramentas de visualização e técnicas de inteligência artificial (IA). Assim, por ser uma ciência multidisciplinar, as inovações elaboradas nestas áreas são directamente reflectidas nas metodologias, técnicas e ferramentas usadas na descoberta de conhecimentos em bases de dados (Oliveira, 2000).



## 1.2 - OBJECTIVOS

O objectivo desta dissertação consiste na extracção de conhecimento de bases de dados relativos às propriedades físicas e químicas de fibras de algodão, provenientes de diversas origens, com a aplicação da ferramenta de prospecção de dados (*Data Mining*) *Clementine*. A aplicação de algoritmos na obtenção de regras capazes de determinar relações entre as propriedades físicas e químicas de fibras de algodão de diferentes origens, bem como o consequente estudo da possibilidade de existência de qualquer relação entre elas, possibilita a inferição das suas propriedades químicas através das suas características físicas. Esta dissertação utiliza e compara o desempenho de diversas técnicas de *Data Mining* aplicadas ao problema de selecção de fibras de algodão de melhor qualidade, para utilização na industria têxtil. São usadas técnicas de aprendizagem supervisionada, tais como: redes neuronais, árvores de decisão e regressão.

A descoberta de conhecimento em bases de dados está relacionada com a identificação de relacionamentos implícitos existentes nos dados estudados. O processo de descoberta de conhecimento, que passa por várias etapas, inclui a gestão dos algoritmos de prospecção de dados, utilizados para extrair padrões desses mesmos dados e a interpretação dos padrões encontrados pelos mesmos.

O objectivo final é criar um modelo de selecção da matéria-prima de forma a classificá-la conforme a qualidade que lhe é inerente, ao nível das propriedades físicas e químicas em estudo, inferindo as possíveis relações entre ambas.

## 1.3 - MOTIVAÇÃO

As tecnologias de sistemas de informação, vem ao longo dos anos a sofrer constantes evoluções no que respeita ao acesso e posterior análise dos dados.

Todos os dias, em qualquer organização, flui uma enorme quantidade de dados. Estes dados contém, e frequentemente escondem, valiosa informação, quer actual, quer histórica, de uma organização. Assim é de interesse das organizações o uso de técnicas de ECBD na tentativa da procura de conhecimento nesta nova realidade.

A aplicação de técnicas de prospecção de dados torna possível comprovar o pressuposto da transformação de dados em informação e posteriormente em conhecimento. Esta possibilidade configura estas técnicas como imprescindíveis para o processo de tomada de decisão. Para se chegar



a este resultado, é preciso investigar o uso efectivo do conhecimento obtido através da prospecção de dados no processo de tomada de decisão, bem como os impactos que teve na solução efectiva de problemas e acções propostas e concretizadas.

Um dos factores que levou à realização deste trabalho deve-se ao facto de em estudos anteriores (Vasconcelos, 1986), se ter verificado que a matéria-prima representa uma parte substancial no custo de produção numa fição. Este facto, por si só, é suficiente para indicar a importância da matéria-prima, quer para o fiandeiro, quer para o utilizador final. As propriedades das fibras de algodão determinam a sua utilização e são de maior importância para os produtores, fiandeiros, acabadores e utilizadores.

O foco do estudo consiste então na utilização do *Clementine*, na busca e determinação das possíveis relações entre as diferentes propriedades físicas da fibra de algodão e as suas características químicas através da aplicação de técnicas de modelação e de aprendizagem máquina existentes na ferramenta. Desta forma obtem-se um fio com as características pretendidas, partindo da utilização da matéria-prima apropriada (fibra).

#### **1.4 - ESTRUTURA DA DISSERTAÇÃO**

A dissertação tem a seguinte organização:

Os capítulos 2 e 3 pretendem fazer um levantamento bibliográfico sobre os vários temas acessórios ao trabalho com o objectivo de recolher opiniões e conclusões de autores avalizados na matéria e simultaneamente permitir ao leitor um mais fácil entendimento de alguns dos termos tratados nesta dissertação. Inegável o contributo deste capítulo para a actualização do autor sobre as matérias focadas nele.

O capítulo 4 descreve a ferramenta utilizada, proporcionando uma descrição do seu interface, bem como de todas as metodologias utilizadas para a obtenção de resultados.

O capítulo 5 expressa em detalhe o trabalho realizado bem como os resultados obtidos com o estudo retirando ilações intermédias de acordo com os resultados parciais.



No capítulo 6, são tiradas conclusões relativas ao trabalho desenvolvido e são estabelecidas ligações que permitirão continuá-lo no futuro, com o objectivo de contribuir para o melhoramento do desempenho da ferramenta utilizada, bem como do processo produtivo onde a fibra de algodão surge como matéria-prima.



# Cap.2 A EXTRACÇÃO DE CONHECIMENTO NAS BASES DE DADOS

---

Hoje em dia, a quantidade de dados armazenados, excede a capacidade de análise desses mesmos dados sem o uso de técnicas de análise automatizadas. Muitas bases de dados comerciais, transaccionais e científicas crescem em grande proporção. A extracção de conhecimento nas bases de dados é um campo que está em evolução por forma a fornecer soluções de análise automatizadas.

Diversos tipos de dados estão a ser recolhidos e acumulados a um ritmo elevado. Surge então a urgente necessidade de uma nova geração de técnicas e ferramentas computacionais de forma assistir o ser humano na extracção de informação útil, isto é, conhecimento, dos enormes volumes e em constante crescimento de dados. Estas técnicas e ferramentas são desenvolvidas devido à necessidade de extracção de conhecimento em bases de dados (ECBD).

O processo de ECBD é visto como um conjunto de actividades, tendo como finalidade a atribuição de significado aos dados armazenados e analisados. Como parte central deste processo surge a aplicação de técnicas de prospecção de dados (*data mining*) para a descoberta de padrões de conhecimento.

A ideia geral de ECBD torna-se tanto atraente, como intuitiva. Tipicamente toda a atenção é focada na aprendizagem de algoritmos, que possam prover um núcleo de capacidades para generalizar a partir de um largo número de factos específicos para um conjunto de regras úteis de alto nível. Estas técnicas de aprendizagem parecem sustentar o conteúdo científico no processo de extracção de conhecimento em bases de dados.

## 2.1 - VISÃO GERAL

A ideia geral de ECBD, tecnicamente, deslumbra-se como um desafio de carácter significativo, bem como difícil. De forma global, o processo de ECBD é considerado como a extracção não trivial de informação implícita, previamente desconhecida e potencialmente útil, a partir dos dados (Fayyad et al., 1996). Assim, é demonstrada uma distinção clara entre a extracção dos dados e a descoberta de conhecimento. Sob suas convicções, a descoberta de conhecimento examina os resultados tal como



provêm da análise dos dados (processo de extrair tendências ou modelos dos dados), após o qual, de forma cuidada e com precisão, os transforma em informação útil e compreensível. Esta informação não é tipicamente recuperável pelas técnicas normais mas sim através do uso de técnicas IA (Inteligência Artificial) (Fayyad et al., 1996).

O Processo de ECBD é um campo em crescente desenvolvimento. Existem muitas metodologias para a extracção de conhecimento em uso mas no entanto denota-se um baixo desempenho por parte destas (Frawley et al., 1991). Algumas destas técnicas são genéricas, enquanto outras são do domínio específico.

## **2.2 - ANTECEDENTES**

Embora existam muitos enquadramentos relativamente ao processo de ECBD, existem seis elementos comuns e essenciais que qualificam a extracção de conhecimento como uma técnica. Os seguintes pontos consistem nas características básicas que compartilham todas as técnicas de KDD: (Frawley et al., 1991; Fayyad et al., 1996)

- ✓ Todas as diferentes técnicas tratam grandes quantidades de dados;
- ✓ É requerida eficiência devido ao enorme volume de dados;
- ✓ A exactidão é um elemento essencial;
- ✓ Todos os métodos requerem o uso de uma linguagem alto nível;
- ✓ Todas as técnicas utilizam alguma forma de aprendizagem automática;
- ✓ Todas produzem alguns resultados interessantes.

São requeridas grandes quantidades de dados que proporcionem informação suficiente, derivando daí conhecimento implícito. Dado que se requerem grandes quantidades de dados, a eficiência consiste no essencial do processo. A exactidão é requerida de forma a assegurar que a extracção de conhecimento é válida. Os resultados terão que ser apresentados de forma compreensível ao ser humano. Uma das maiores premissas do ECBD é que o conhecimento é descoberto usando técnicas de aprendizagem automática que irão examinar os dados através de processos automatizados. Para estas técnicas serem consideradas úteis, a descoberta do conhecimento deve ser interessante, isto é, deve ter um valor potencial para o utilizador.



### **2.3 - EM QUE CONSISTE A PROSPECÇÃO DE DADOS (DATA MINING)?**

O crescimento rápido de dados, recolhidos e armazenados em enormes e variadas quantidades de bases de dados, excedeu a capacidade humana de compreensão sem o recurso a ferramentas apropriadas. Como resultado, os dados coleccionados em grandes bases de dados transformaram-se em “sepulturas de dados”, isto é, dados raramente consultados. Assim, a prospecção de dados, consiste na forma de extracção de conhecimento válido dessas grandes quantidades de dados.

“A necessidade é a mãe da invenção” (Han et al., 2001).

A principal razão pela qual surgiu uma grande motivação por parte da indústria de tecnologias de informação, nos processos de prospecção de dados, está relacionado com a existência de grandes quantidades de dados e a consequente necessidade de os tornar em informação útil e consequentemente, em conhecimento.

A prospecção de dados pode ser vista como um resultado natural da evolução das tecnologias de informação.

### **2.4 - O ECBD, PROSPECÇÃO DE DADOS E A RELAÇÃO COM OUTROS CAMPOS**

Historicamente, foi dado uma variedade de nomes à acção de descoberta de padrões úteis sobre as bases de dados, nomes do tipo, *data mining*, extracção de conhecimento, descoberta de informação, colheita de informação, pesquisa sobre os dados, processamento de padrões de dados, etc. O termo prospecção de dados ou *data mining*, tem sido mais utilizado por estatísticos, analistas e comunidades de gestores de sistemas de informação (GSI). O termo ECBD foi atribuído no primeiro Workshop sobre ECBD em 1989 (Piatetsky-Shapiro, 1991) de forma a dar ênfase a que o “conhecimento” consiste no produto final da descoberta do significado dos dados. Este termo foi popularizado em processos relacionados com a inteligência artificial e aprendizagem maquina.

O ECBD refere o processo global de descoberta de conhecimento útil a partir de dados, enquanto que a prospecção de dados refere-se a um passo particular dentro deste processo. A prospecção de dados consiste na aplicação de algoritmos específicos por forma a extrair padrões dos dados em análise. Os passos adicionais num processo de ECBD, tipo preparação, pré-processamento dos dados, incorporando conhecimento prévio e uma interpretação própria dos resultados da exploração, são essenciais de forma a assegurarem que, a partir dos dados analisados deriva conhecimento útil.



O estudo da automatização da análise dos dados tem-se desenvolvido, quer sob a denominação de Prospecção de Dados, quer sob a designação mais geral de Extracção de Conhecimento em Bases de Dados (ECBD) e envolve teorias, algoritmos e métodos de várias áreas científicas: Bases de Dados, Matemática, Estatística, Aprendizagem, e Visualização Gráfica entre outras.

Como o processo de ECBD está intimamente relacionado com bases de dados, surge uma outra área relacionada, área esta, denominada de *data warehousing*. Esta área refere-se à preocupação de coleccionar e seleccionar dados transaccionais tornando-os disponíveis para análises *on-line* e suportes de decisão. Uma abordagem popular para análise de *data warehousing* é denominada de OLAP<sup>3</sup> (Codd, 1993). As ferramentas OLAP focam-se na análise de dados multidimensionais. Estas ferramentas são criadas com o objectivo de suportar e simplificar a análise de dados de forma interactiva, enquanto que as ferramentas de ECBD tem como objectivo automatizar quanto possível o processo (Codd, 1993).

## 2.5 - PROSPECÇÃO DE DADOS, DEFINIÇÃO E CONCEITOS GERAIS

O termo “prospecção de dados” é apenas um de diversos termos, que descrevem realmente o conceito relativo à extracção de conhecimento das bases de dados. A ideia suporte ao processo de ECBD surge então como: (Fayyad et al., 1996)

*“O processo não trivial de identificação de padrões de dados validos, novos, potencialmente úteis e compreensíveis”.*

Para compreender melhor o conteúdo desta esparsa definição, deve-se olhar individualmente cada componente da mesma:

- Dados: Conjunto de factos ou casos existentes num repositório de dados;
- Padrão: designa os conjuntos de modelos que representam alguma abstracção de um subconjunto de dados numa qualquer linguagem descritiva;
- Processo: a extracção de conhecimento de base de dados envolve diversas etapas como a preparação dos dados, procura de padrões e avaliação do conhecimento;

---

<sup>3</sup> On-line analytical processing



- Válidos: os padrões descobertos devem possuir algum grau de certeza. Ou seja, devem satisfazer funções ou atingir patamares que garantam que os exemplos cobertos e os casos relacionados com o padrão encontrado sejam aceitáveis;
- Novos: um padrão encontrado deve fornecer novas informações sobre os dados. O grau de novidade serve para determinar o quão novo ou inédito é um padrão. Ele pode ser medido através de comparações entre as mudanças ocorridas nos dados ou no conhecimento existente anteriormente;
- Úteis: os padrões descobertos devem ser incorporados de forma a serem utilizados;
- Compreensíveis: um dos objectivos ao realizar-se um processo de ECBD é fazer com que os padrões possam ser entendidos pelos utilizadores finais através de uma qualquer linguagem descritiva, de forma a permitir uma compreensão mais profunda dos dados;
- Conhecimento: o conhecimento está intimamente relacionado com o domínio de estudo, estando este fortemente relacionado com medidas de utilidade, originalidade e compreensão.

É de realçar que vários autores usam o termo prospecção de dados para referir o processo de extracção de conhecimento em bases de dados, porém, os termos ECBD e prospecção de dados são usados com significados distintos. ECBD refere-se ao processo como um todo, enquanto que, prospecção de dados é tratado como um componente dentro deste processo.

## **2.6 - PROCESSO DE ECBD**

Extrair conhecimento de bases de dados pode envolver, entre outras coisas, a utilização de algoritmos de aprendizagem máquina capazes de generalizar os factos encontrados num grande repositório de dados na forma de regras de alto nível, compreensíveis ao ser humano (Brachman et al., 1996).

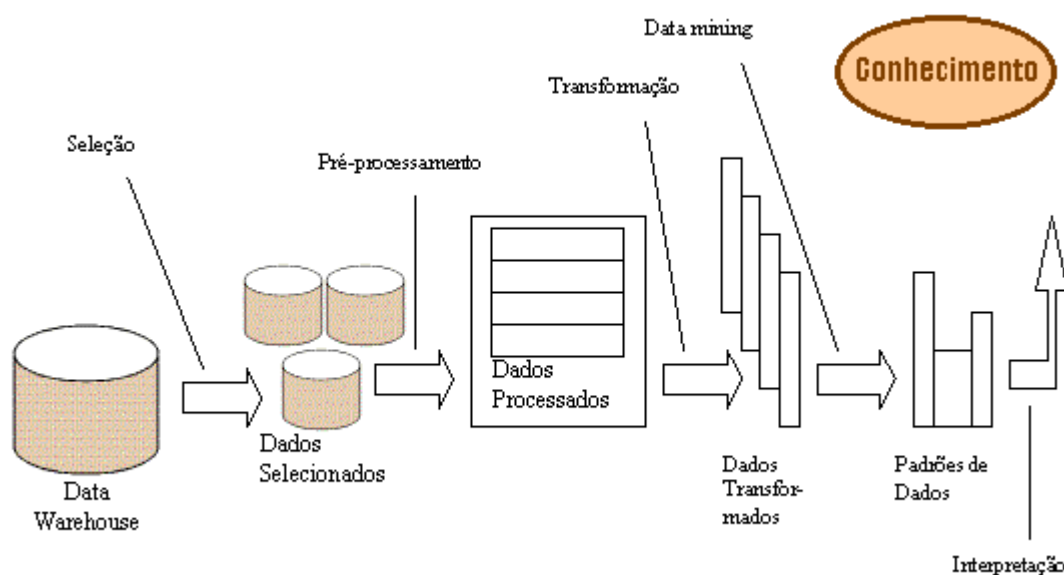
As técnicas de aprendizagem máquina parecem estimular bastante os investigadores da área e talvez seja a parte do processo de ECBD que contenha, substancialmente, os maiores conteúdos científicos estudados. Todavia, quando se tenta realizar tarefas de extracção de conhecimento, em aplicações do mundo real, percebe-se que as mesmas podem ser extremamente complexas e que a tarefa de



prospecção de dados representa apenas uma porção (importante) do processo global (Brachman et al., 1996).

O processo de extracção de conhecimento em bases de dados é interactivo e iterativo, envolvendo inúmeros passos com várias decisões a serem tomadas pelo utilizador. Brachaman & Armand (Brachman et al., 1996), dão uma visão prática e sobre todo o processo de ECBD, dando ênfase à natureza interactiva do processo (figura 1).

Assim, o processo de ECBD deve ser visto, como sendo composto por várias etapas interligadas entre si.



**Figura 1** – Visão generalista do processo de ECBD (Abonyi et al., 2000).

O processo de extracção de conhecimento em bases de dados é interactivo, envolvendo diversas etapas, como é demonstrado na figura 1, no entanto esta visão é um pouco generalista.

Assim, olhando de forma menos generalista todo o processo de ECBD, podemos deslumbrar este mesmo processo de uma forma um pouco distinta e mais aprofundada do que está descrito e visualizado na figura 1.



Assim sendo, podemos considerar os seguintes passos de forma a efectuar a extracção de conhecimento (Fayyad et al., 1996):

1. Compreensão do domínio de aplicação;
2. Selecção de dados;
3. Pré-processamento dos dados;
4. *Data mining*;
5. Interpretação e validação do conhecimento;
6. Consolidação do conhecimento.

As etapas do processo de ECBD, normalmente, são realizadas de forma sequencial, ou seja, é necessário compreender o domínio de aplicação, seleccionar e pré-processar os dados por forma a se encontrar padrões nesses mesmos dados. E por se tratar de um processo interactivo, as pessoas envolvidas na sua realização podem possuir um canal de comunicação que viabilize uma boa troca de informações. Estes utilizadores possuem diferentes especialidades e necessidades e cada um possui o seu papel dentro do processo. As classes dos usuários envolvidos no processo de ECBD são (Brachman et al., 1996):

- Especialista do domínio: possui ampla compreensão do domínio da aplicação. Por exemplo, durante a análise de um conjunto de dados sobre matéria-prima de uma organização, o especialista poderá ser o gestor de stocks.
- Analista: executa o processo de ECBD e, portanto, deve possuir um amplo conhecimento das etapas que compõem esse processo. Por exemplo, analistas de sistemas, estatísticos e administradores de bases de dados.
- Utilizador final: pode ou não possuir amplo conhecimento do domínio. É frequentemente aquele que utiliza o conhecimento extraído no processo de ECBD para auxiliá-lo num processo de tomada de decisão.

Durante a realização do processo de ECBD, a interacção e comunicação entre o analista, o especialista do domínio e o utilizador final é muito importante, uma vez que o sucesso da extracção de conhecimento depende da comunicação entre eles (Brachman et al., 1996). Nesse sentido, é possível melhorar a troca de informações entre os envolvidos no processo, utilizando-se técnicas de



visualização de dados adequados para auxiliar a execução das diferentes etapas do processo (Rezende et al., 1998b).

As etapas do processo de ECBD sugeridas podem não cobrir exhaustivamente todos os passos existentes, entretanto, esta sequência de tarefas pode ser considerada como consistindo nos requisitos mínimos de forma a se conseguir extrair conhecimento de bases de dados. Todas as etapas são estruturadas e detalhadas seguidamente.

### **2.6.1 - Compreensão do Domínio de Aplicação**

Um aspecto importante no processo de ECBD consiste na compreensão do domínio de aplicação, de onde será extraído o conhecimento necessário para apoiar a tomada de decisões. Torna-se evidente a necessidade de identificar quais são os objectivos e requisitos necessários para que seja possível extrair o conhecimento possível (Rocha, 1999).

Não existem perguntas padrão cujas respostas possam delinear um domínio de aplicação, uma vez que cada situação mantém as suas peculiaridades próprias. Porém, alguns aspectos podem ser levados em consideração na tentativa de se definir a esfera de acção de uma aplicação, tais como:

- Definição dos objectivos e dos requisitos da extracção de conhecimento;
- Identificação das possíveis fontes de dados;
- Pesquisa do conhecimento prévio existente;
- Modo de distribuição do conhecimento extraído;
- Incorporação do conhecimento extraído num sistema de tomada de decisão;
- Estudo de viabilidade e custos da aplicação do processo.

Uma vez que o domínio de aplicação tenha sido compreendido, deve-se seleccionar e preparar os dados para serem analisados numa próxima etapa.



### 2.6.2 - Selecção do Dados

Com o domínio especificado, deve-se localizar e escolher quais as fontes de dados que estão relacionadas com este, para que um conjunto de dados apropriado possa ser seleccionado e retirado das mesmas (Rocha, 1999).

As fontes de dados utilizadas para realizar o processo de ECBD podem ser classificadas como internas e externas. As fontes internas normalmente são repositórios de dados que já estão incorporados no sistema de aplicação do domínio em questão, e normalmente, são constituídas por *data warehouses*, bases de dados operacionais e dados históricos. As fontes externas são compostas por todos os outros tipos de dados que habitualmente não estão incorporados ao sistema de aplicação, como por exemplo, documentos (e.g. relatórios), informações do especialista do domínio, etc.

Os dados existentes em fontes internas normalmente já estão armazenados em algum tipo de repositório estruturado (e.g. bases de dados relacionais) e por isso, requerem menos processamento, já que eles podem ser recuperados utilizando-se, por exemplo, consultas SQL<sup>4</sup>. No entanto os dados contidos em fontes externas necessitam formatação e de serem inseridos no sistema para que possam ser recuperados e preparados para análise (Rocha, 1999).

É interessante citar que, frequentemente, os dados de análise podem nem existir e devem ser armazenados assim que os eventos vão acontecendo.

As fontes podem possuir quantidades de *gigabytes* ou *terabytes* de dados e, além disso, estes dados podem ser provenientes de fontes distintas, necessitando que estes sejam reduzidos e colocados num conjunto de dados com estrutura homogénea (e.g. tabela) para permitir a análise conjunta destes.

Os dados podem estar armazenados em diferentes sítios, exigindo que os mesmos sejam integrados. Esta tarefa de integração não é trivial, exigindo a adequação de alguns dados (Fayyad et al., 1996).

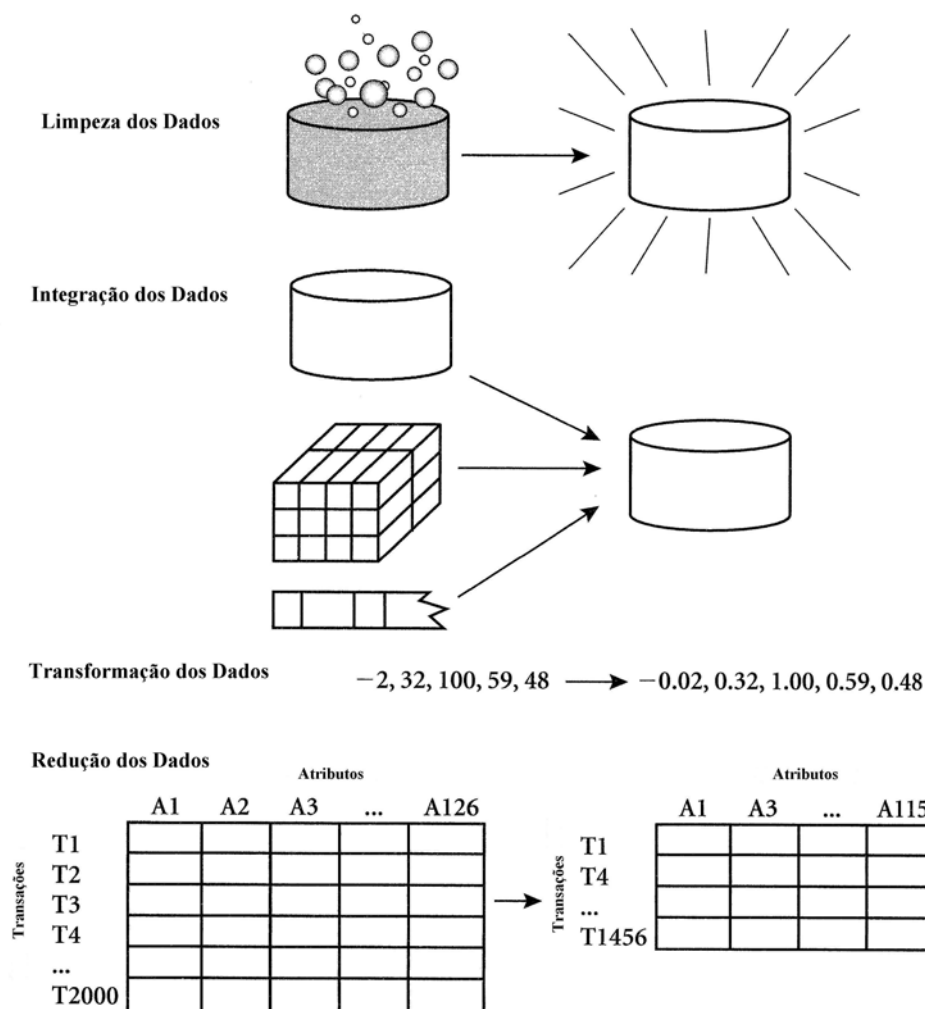
Seguidamente é descrita a etapa responsável pela resolução desses tipos de problemas.

### 2.6.3 - Pré-processamento dos dados

A figura 2 demonstra as etapas referentes ao pré-processamento dos dados.

---

<sup>4</sup> Structured Query Language



**Figura 2** – Etapas constituintes do pré-processamento dos dados (Han et al., 2001).

No pré-processamento dos dados, normalmente, é realizada a amostragem de dados. Por vezes nem sempre é possível analisar todos os dados de uma população, sendo necessário retirar uma amostra representativa para realizar o processo de ECBD. Os dados também podem estar em formatos distintos, pois os mesmos poderão ter sido recolhidos de diferentes fontes. Assim, algumas transformações devem ser efectuadas, por exemplo, converter tipos de dados não suportados num outro tipo em que todas as fontes consigam representar, se isso for possível (Rocha, 1999).

Outra problemática que deve ser resolvida envolve o tamanho do conjunto de dados, pois a grande quantidade de dados reunida, às vezes, podem impossibilitar a realização do processo de ECBD uma vez que alguns algoritmos usados na prospecção de dados conseguem tratar apenas um número limitado de registos. Devido a isso, devem-se utilizar algumas técnicas de amostragem para que se



possa reduzir o tamanho do conjunto de dados, obtendo um subconjunto que seja relevante e representativo.

A realização de amostragem é crítica dentro do processo de ECBD, pois se não forem utilizadas técnicas estatísticas adequadas, pode-se gerar um subconjunto de dados não representativo, resultando análises que não demonstram a verdadeira situação dos factos registados na base de dados ou distorcendo os possíveis resultados. Deste modo, devem ser observados os seguintes aspectos para se efectuar uma amostragem dos dados, tais como:

- O tamanho da amostra;
- Estratégias para a obtenção da amostra (técnicas estatísticas que podem ser utilizadas, bem como o apoio do especialista do domínio);
- Homogeneidade dos dados;
- Dinâmica dos dados (mudanças de valores de atributos ao longo do tempo).

Após reunir dados de diversas fontes, é muito provável que o conjunto de dados venha a ter registos duplicados, ruídos (erros nos dados) e muitos dados ausentes. Mesmo que a recolha de dados seja efectuada de apenas uma fonte, estes tipos de problemas podem ainda ocorrer. Por isso, deve-se efectuar uma integração e posterior transformação dos dados (Rocha, 1999).

As operações de remoção de ruídos dos dados, tratamento de atributos ausentes e remoção de registos repetidos efectuados nesta fase compreendem, entre outros, os seguintes aspectos:

- Padronização dos valores dos atributos: num conjunto de dados constituído por diversas fontes pode acontecer, por exemplo, que o campo “Origem” possua diferentes valores e tipos com o mesmo significado como, “Origem”, “Orig”, “Org”. Então, deve-se transformar estes valores como sendo de um tipo único e iguais para todos;
- Remoção de registos duplicados: pode ocorrer que os mesmos dados de uma entidade em estudo possam estar contidos em diferentes registos e sendo considerado como dois elementos distintos, pois, por exemplo a designação de uma entidade num determinado registo pode aparecer sem abreviações e em outro aparece abreviado.
- Eliminação de ruídos: muitas vezes os dados colectados podem conter erros ocasionados por diversos factores no momento da recuperação de diversas fontes, (e.g. erros de digitação da fonte original, sendo esta em papel). Os campos que contém ruído nos seus



valores devem ser tratados atribuindo o valor correcto aos dados ou devem ser simplesmente eliminados da base de dados, caso não tenham forma de ser tratados.

- Valores ausentes: é muito comum encontrar registos cujos campos possuam valores ausentes. Isto pode ocorrer, entre outras coisas, devido a erros na introdução dos dados. Desta forma deve-se estabelecer critérios para o tratamento de atributos ausentes (Rocha, 1999).

Por fim, as técnicas de redução dos dados podem ser utilizadas por forma a obter uma representação mais reduzida da base de dados em estudo, reduzindo assim o seu volume, mantendo a integridade da base de dados original. Isto é, a extracção de conhecimento em bases de dados assentada num estudo em bases de dados mais reduzidas torna-se mais eficiente produzindo os mesmos resultados analíticos (Han et al., 2001).

Toda esta etapa deve ser feita, sempre que possível, com o acompanhamento do especialista do domínio, pois como este possui um conhecimento mais abrangente do domínio em questão, ele poderá apoiar a decisão acerca dos atributos que podem ser descartados ou que devem ser mantidos. Apesar de existirem ferramentas para a execução desta etapa, a participação de um especialista pode torna-la muito mais precisa e fiável.

Segundo Mannila (Mannila, 1996) as etapas apresentadas até este ponto podem corresponder a 80% de todo o tempo gasto durante o processo de ECBD.

#### **2.6.4 - Data Mining ou Prospecção de Dados**

A etapa de prospecção de dados pode ser vista como uma sequência de etapas dentro do processo de KDD, que pode envolver repetidas iterações da aplicação de um determinado algoritmo com o objectivo de extrair padrões dos dados. É importante frisar que o termo padrão deve ser entendido como sendo uma instanciação de um determinado modelo (Fayyad et al., 1996; Mitchel, 1997).

Os algoritmos e as técnicas usados de forma a gerar modelos a partir de dados, normalmente, provêm de áreas como a aprendizagem máquina, reconhecimento de padrões e a estatística. Estas técnicas, muitas vezes, podem ser combinadas por forma a obter melhores resultados.



#### 2.6.4.1 - FUNÇÕES DA PROSPECÇÃO DE DADOS

Os dois objectivos principais da prospecção de dados são a previsão e a definição de modelos.

A previsão pode ser conseguida utilizando-se determinadas variáveis para prever valores desconhecidos ou futuros de outras variáveis de interesse. A definição envolve a descoberta de padrões interpretáveis por humanos que descrevam os factos registados em bases de dados. Alguns autores (Fayyad et al., 1996) mencionam que dentro do contexto de ECBD, definir modelos possui maior importância que prever os mesmos, embora a maioria dos algoritmos de aprendizagem máquina e reconhecimento de padrões foquem a previsão.

A previsão e a definição podem ser realizadas através das seguintes tarefas:

- **Classificação:** a classificação é uma tarefa predictiva que pode ser usada para encontrar um modelo que classifique um registo entre várias classes previamente definidas. Corresponde a uma função de em que cada caso é classificado numa classe de entre um conjunto de classes pré-definidas;
- **Regras de Associação:** responsável por encontrar relações entre os campos de um conjunto de dados;
- **Regressão:** tarefa predictiva cujo objectivo é encontrar uma função de mapeamento para um registo, normalmente numérico, que possibilite a previsão dos valores do mesmo;
- **Segmentação:** tarefa descritiva cujo objectivo é procurar identificar um conjunto de categorias finitas ou agrupamentos naturais que possam descrever um certo comportamento nos dados. Estas categorias podem ser mutuamente exclusivas ou podem consistir numa representação hierárquica;
- **Resumo ou Síntese:** tarefa que visa obter uma descrição compacta de um conjunto de dados para se conseguir fazer análises exploratórias de dados e geração automática de relatórios;
- **Modelação de dependência:** Consiste na procura de um modelo que descreva dependências significativas entre variáveis; estes modelos de dependências existem a dois níveis: o nível de estrutura em que o modelo especifica que variáveis estão localmente dependentes, e o nível quantitativo que especifica a intensidade das dependências numa escala numérica;



- Evolução: esta tarefa está focada na descoberta das mudanças mais significativas nos dados através de medidas anteriores ou de valores normativos. O objectivo é modelar o processo gerando uma sequência ou relatando tendências do processo ao longo do tempo.

Com a definição da tarefa referente à prospecção de dados, pode ser escolhido o algoritmo ou uma combinação deles para serem utilizados. De seguida são apresentados alguns dos principais algoritmos.

#### 2.6.4.2 - ALGORITMOS USADOS NA PROSPECÇÃO DE DADOS

A escolha do algoritmo para a prospecção de dados não é uma tarefa trivial, pois, deve-se considerar vários aspectos, tais como, o objectivo da tarefa a ser realizada, o tamanho da base de dados, a precisão exigida, entre outros. Muitas vezes, são realizados testes empíricos com diferentes tipos de algoritmos para se descobrir qual deles apresenta o melhor resultado em termos da qualidade e precisão dos padrões encontrados.

Para que o analista possa seleccionar qual o algoritmo que melhor se enquadra a uma determinada tarefa, este deve conhecer quais os aspectos básicos de cada potencial método. Os algoritmos de aprendizagem máquina podem ser classificados em vários aspectos: por tipo de aprendizagem, por tipo de paradigma, por tipo de linguagem de descrição e por modo de incorporação de novos exemplos. A tabela 1 apresenta um resumo desta classificação.

Tipo de Aprendizagem	Paradigmas de Aprendizagem	Linguagem de Descrição	Modo de incorporação de novos exemplos
Supervisionado	Simbólico	Exemplo ou Instancias	Incremental
Não Supervisionado	Estatístico	Hipóteses ou Conceitos Aprendidos	Não Incremental
	<i>Instance-Based</i>	Teoria do domínio ou Conhecimento de Fundo	
	Conexionista		
	Genético		

**Tabela 1** – Características gerais dos sistemas de aprendizagem máquina (Kohavic et al., 1994).



Com relação ao modo como estes algoritmos aprendem, eles podem ser classificados como sendo de aprendizagem supervisionada e não supervisionada. Na aprendizagem supervisionada os algoritmos utilizam dados para treino e para validação e posterior teste dos modelos gerados. Na aprendizagem não supervisionada não existem dados de treino, assim, o algoritmo deve tentar encontrar os padrões através da caracterização e/ou segmentação destes dados.

Os algoritmos de aprendizagem máquina podem ser de diferentes paradigmas, tais como:

- Paradigma simbólico: a aprendizagem é feita na tentativa de se construir representações de um conceito através da análise de exemplos e contra-exemplos. As apresentações simbólicas estão tipicamente na forma de alguma expressão lógica, árvore de decisão, rede semântica, etc. Os métodos de aprendizagem simbólico podem ser tratados como: proposicional e relacional (Tecuci et al., 1995).
- Paradigma Estatístico: as técnicas estatísticas, em geral, tendem focar tarefas em que todos os atributos têm valores contínuos. Vários modelos matemáticos são utilizados para se construir os modelos de dados (Elder et al., 1995).
- Paradigma *instance-based*: a classificação de um caso é efectuada através do conhecimento de um caso anterior e o assumir que o novo caso terá as mesmas características. Esta filosofia exemplifica os sistemas *instance-based*, que classificam casos nunca vistos através de casos similares conhecidos (Quinlan, 1993).
- Paradigma Conexionista: as RNA<sup>5</sup> são construções matemáticas relativamente simples que utilizam o mecanismo de paralelismo, onde são conectados um grande número de pequenas unidades de processamento ligadas em rede. As RNA possuem a capacidade de aprender através de exemplos e efectuar interpolações e extrapolações da aprendizagem efectuada. Neste tipo de paradigma não se procura obter regras como na abordagem simbólica, mas determinar a intensidade de conexões entre neurónios (Braga et al., 2000).
- Paradigma Genético: um classificador genético consiste numa população de elementos de classificação que competem de forma a efectuar um vaticínio, onde os elementos que possuem uma performance fraca são descartados e os mais fortes proliferam, produzindo variações de si mesmos. Os algoritmos deste paradigma são conhecidos por

---

<sup>5</sup> Redes neuronais artificiais



algoritmos genéticos, os quais se baseiam nos mecanismos de selecção natural e genético por forma a efectuar optimizações (Goldberg, 1998).

Num sistema de aprendizagem é necessário representar exemplos, conceitos ou a teoria do domínio da aplicação na forma de linguagens de descrição. Estas linguagens consistem em (Oliveira, 2000):

- Linguagem de descrição de exemplos ou instâncias: descreve os exemplos utilizados pelo programa para aprender conceitos, estabelecendo limites sobre os tipos de padrões que o sistema pode aprender.
- Linguagem de descrição de hipóteses ou conceitos aprendidos: descreve o estado interno de um programa de aprendizagem, correspondente à teoria dos conceitos ou padrões existentes nos dados, estabelecendo limites sobre o que pode ou não ser aprendido (e.g. árvores de decisão).
- Linguagem de descrição da teoria de domínio ou conhecimento de fundo: descreve todo o conhecimento prévio que o programa possui a respeito do domínio.

Os algoritmos de aprendizagem indutivos podem ser classificados de acordo com a possibilidade de integração de novos exemplos aos modelos encontrados, podendo ser classificados como (Oliveira, 2000):

- Não Incremental: necessita que todos os exemplos de treino, em simultâneo, estejam disponíveis para que seja induzido um conceito. É vantajoso usar estes algoritmos para problemas de aprendizagem onde todos os exemplos estão disponíveis e provavelmente não irão ocorrer mudanças.
- Incremental: revê a definição do conceito corrente, se necessário, em resposta a cada nova instância de treino observada. Os exemplos observados são considerados um a um pelo sistema, isto é, o sistema considera o primeiro exemplo e, de acordo com esse exemplo, constrói uma determinada hipótese; a seguir considera um segundo exemplo, que pode ou não modificar a primeira hipótese, baseando-se em como esta classifica o segundo exemplo.



### 2.6.4.3 - EXTRACÇÃO DE PADRÕES

Uma vez definidas as funções da prospecção de dados e o algoritmo escolhido, pode-se efectuar a extracção de conhecimento. Entretanto, deve ser tido em conta que o analista necessita de observar alguns aspectos relativos à extracção de conhecimento, tais como (Oliveira, 2000):

- Definição de parâmetros: algoritmos diferentes, normalmente, possuem parâmetros distintos. A definição correcta dos parâmetros, frequentemente, pode ser conseguida através de experiências empíricas. Ou seja, quanto mais experiente for o analista maior serão as hipóteses deste conseguir ajustar os parâmetros adequadamente.
- Redução de atributos: pode ser reconhecido que alguns dos atributos do conjunto de dados não possuem relevância para o algoritmo seleccionado (e.g. atributos numéricos para algoritmos simbólicos), então, pode-se pensar na exclusão de tais atributos.
- Generalização: um algoritmo deve ser capaz de reconhecer padrões desconhecidos e não decorar exemplos apresentados. Para isso, deve-se utilizar técnicas estatísticas que reforçam a generalização de um preditor ou descritor, além de garantir pequenas taxas de erros. Para isso, as amostras são divididas em conjuntos separados de treino e validação. Possibilitando, desta forma, medir a taxa de erro através do conjunto de teste. Estas técnicas são (Oliveira, 2000):
  - *Holdout*: este método divide fixamente a amostra de exemplos em 2/3 para treino e o restante 1/3 para validação. A estimativa deste método é adequada a grandes conjuntos de dados.
  - *Resampling*: estes métodos realizam experiências com diferentes partições das instâncias em teste e validação e possui como taxa de erro a média das taxas de erros calculadas sobre o conjunto de teste em cada experiência. Um método bastante utilizado de *resampling* é o *cross-validation* que divide um conjunto de exemplos em  $k$  partições mutuamente exclusivas. A cada iteração, é utilizada 1 partição para testar o sistema de aprendizagem e as  $k-1$  restantes para a validação. Como são realizadas  $k$  iterações, então todas as  $k$  partições servem como conjunto de validação.
- Mudanças dos dados: dados que mudam com muita frequência podem invalidar os padrões descobertos anteriormente. A possível solução para este tipo de problema consiste em incluir métodos incrementais que consigam actualizar os padrões já encontrados.



- Compreensão de padrões: em muitas aplicações é importante tornar as descobertas mais facilmente compreensíveis ao utilizador final. Assim, podem ser utilizadas técnicas sofisticadas de visualização dos dados. Os resultados obtidos, normalmente, são apresentados em algum tipo de representação como árvores de decisão, regras de decisão, gráficos ou relatórios escritos.

Várias áreas estão envolvidas no processo de ECBD, desde a recolha dos dados a serem analisados até à visualização dos resultados obtidos.

### **2.6.5 - Interpretação e Avaliação do Conhecimento**

Os padrões encontrados devem ser validados a partir da interpretação e avaliação destes. Os utilizadores envolvidos devem interpretar os padrões extraídos, e para isso, podem lançar mão a ferramentas estatísticas e de visualização que permitam efectuar uma “leitura” precisa dos resultados obtidos, de forma a possibilitar a verificação da validade e novidade, ou mesmo, a relevância dos padrões encontrados.

É importante notar que esta tarefa deve ser feita em conjunto pelos utilizadores envolvidos no processo de extracção de conhecimento, pois somente com o auxílio de especialistas é que o analista terá condições de avaliar se o que foi desconcerto é ou não um conhecimento relevante.

Caso o conhecimento não seja validado, então, provavelmente, deve-se retornar às etapas anteriores e tentar refazê-las ou melhora-las. Esta iteração pode ocorrer até que se obtenha resultados aceitáveis ou concluir-se que não seja possível extrair conhecimento relevante dos dados em análise.

### **2.6.6- Consolidação do Conhecimento**

O conhecimento depois de validado deve ser consolidado, ou seja, deve associar valor a um sistema de apoio à tomada de decisão. Isto é feito incorporando-se tal conhecimento a um sistema de aplicação ou, simplesmente, documentando-o e relatando-o às partes interessadas. Após a consolidação do conhecimento, pode-se resolver conflitos potenciais entre os conhecimentos anteriores, previamente extraídos e reconhecidos.



## **2.7 - CONSIDERAÇÕES FINAIS**

Sistemas capazes de auxiliar a aquisição de conhecimento por forma a solucionar problemas existentes, consistem nos produtos desejados por parte dos gestores de sistemas de informação (Wiederhold, 1996).

A ECBD está em crescimento e com o compromisso de se tornar rapidamente num grande campo de aplicação. A extracção de conhecimento pretende ser uma nova tecnologia no campo das bases de dados nos próximos anos. A necessidade da existência de ferramentas para a extracção automatizada causou o aparecimento de inúmeras ferramentas disponíveis no mercado.

Várias técnicas são usadas na tentativa de extrair conhecimento de factos. É possível extrair conhecimento de especialistas ou através dos dados. A extracção de conhecimento através de especialistas (explícita) visa sintetizar o conhecimento de um especialista humano modelando-o para que este seja incorporado a um sistema. Já a extracção de conhecimento de dados (implícita) é uma forma automática de se extrair conhecimento a partir de dados mantidos em alguma estrutura de armazenamento utilizando-se técnicas e ferramentas para encontrar modelos e padrões.

A extracção de conhecimento de bases de dados é uma área da ciência de computação que visa extrair padrões e modelos de grandes quantidades de dados. Os padrões podem ser extraídos realizando-se passos, como a selecção, pré-processamento, prospecção de dados, avaliação e consolidação do conhecimento.

É antecipado dizer que os sistemas de bases de dados no futuro incluirão capacidades de ECBD, sob a forma de interfaces com bases de dados inteligentes. Alguns tipos de recuperação de informação poderão tirar benefício do uso de técnicas de ECBD. Devido à potencial aplicação da extracção de conhecimento, em diversas áreas denota-se um crescimento de oportunidades de investigação neste campo (Oliveira, 2000).

Como qualquer área em pleno desenvolvimento, novos desafios devem ser continuamente superados. No contexto deste trabalho, é focado todo o processo de ECBD bem como a ferramenta utilizada para o processo de prospecção de dados.



# Cap. 3

## A FIBRA DE ALGODÃO

---

### 3.1 - NOTA HISTÓRICA

A utilização de algodão para a produção de vestuário perde-se nos tempos da Antiguidade. É provável que a fiação e a tecelagem do algodão tenham sido originárias da Índia - os poetas dedicaram-lhe versos 1500 anos antes de Cristo - embora o Egipto reclame ter sido o primeiro a produzir roupas a partir do algodão, assim como do linho. Dentro dos túmulos, foram encontrados tecidos de algodão a envolver as múmias com mais de 4000 anos (Vasconcelos, 1993).

O mais antigo engenho que se conhece para descaroçar o algodão é a “ Chuarca “, na Índia, onde desde os tempos mais remotos, existiam engenhos rudimentares de fiar e tecer. A indústria têxtil do algodão esteve monopolizada pela Índia durante cerca de 3000 anos, tendo atingido tal perfeição que alguns dos seus produtos, como as musselinas, eram tão finos que os outros países nunca conseguiram imitá-los, a não ser depois da segunda metade do século XIX (Queiroz, 1965).

Ao longo dos séculos, foi crescendo a utilização do algodão para fins têxteis, graças à divulgação dos artigos de algodão por navegadores, exploradores e comerciantes. Nessa altura o algodão chegava à Europa, trazido pelas caravanas vindas do Oriente que o vendiam aos mercadores de Génova e Veneza. Em Portugal , em meados do século XIII, no reinado de D. Sancho II, a fibra já era importada.

Contudo não é só a Oriente que a fibra é cultivada. No século XV, com a expansão marítima portuguesa pela costa Africana, descobre-se que o algodão já era cultivado na Guiné e sul de África desde tempos remotos. Também os navegadores espanhóis como Colombo e Cortez, o encontraram nas Índias Ocidentais e México, assim como os portugueses, no Brasil.

Crê-se terem sido os portugueses, o primeiro país europeu a negociar e a utilizar a fibra de algodão, recebida por via marítima.

O primeiro registo do comércio de algodão em Inglaterra data de finais do século XV. A partir da segunda metade do século XVI, tecelões flamengos fixaram-se em Inglaterra, particularmente em Lancashire, onde desenvolveram a arte da manufactura de artigos de lã e algodão, cuja produção cresceu muito lentamente durante o século XVII.



Finalmente com a Revolução Industrial e a invenção de maquinaria têxtil, a produção e a qualidade dos artigos de algodão aumentou gradualmente. Durante a primeira metade do século XIX, o número de invenções continuou a aumentar, se bem que muitas delas consistiam em modificações e melhoramentos das já existentes. Da segunda metade do século XIX até à Segunda Guerra Mundial, não houve grandes mudanças nas máquinas de fiar e tecer o algodão.

### 3.2 - MORFOLOGIA DA PLANTA

O algodão é a penugem que envolve as sementes do algodoeiro – planta do género *Gossypium* – que possui quatro variedades que dão fibras utilizáveis na indústria têxtil. Num passado distante, estas quatro variedades foram cultivadas independentemente por povos tropicais, duas delas no hemisfério Oriental e duas no Ocidental. As espécies e as civilizações ancestrais associadas ao seu cultivo são (Hamby, 1965):

- *Gossypium Arboreum*: planta vivaz com 2 a 7 metros de altura - Índia;
- *Gossypium Herbaceum*: arbusto com 0,5 a 1,5 metros de altura - não existem certezas quanto à sua origem, mas está associada à expansão sob a influência árabe;
- *Gossypium Hirsutum*: arbusto com 0,8 a 2,5 metros de altura - México, América Central e Índias Ocidentais;
- *Gossypium Barbardense*: arbusto com 0,8 a 2,0 metros de altura - Perú.

As duas primeiras variedades são designadas como “ Tipo Velho Mundo “, são produzidas na Ásia – Índia, Paquistão, China, Irão, Iraque, Turquia e Rússia; na África – Egipto, Sudão e países sub-saarianos; e na Europa – Grécia e Espanha. As suas fibras são curtas e grossas (Queiroz, 1965).

O *Gossypium Arboreum*, a “árvore“ originária dos algodões da Índia e da África, era tradicionalmente um arbusto de grande porte. É agora considerado de menor importância comercial.

As outras duas variedades são designadas como “ Tipo Novo Mundo “ são produzidas na América do Norte – E.U.A e na América do Sul – Peru, Brasil e Argentina. As suas fibras são mais finas e longas que as primeiras.



Os algodões Sea Island, Egipcíos e Egipcíos–Americanos pertencem a este grupo e todos têm fibras finas, compridas e sedosas.

Em qualquer das variedades o caule é lenhoso e erecto com os ramos maiores que saem da sua base e vão decrescendo sucessivamente em comprimento até ao cimo, pelo que a configuração da planta é cilíndrica ou cónica.

As folhas prendem-se ao caule por ordem alternada, formando à sua volta espirais regulares e conforme as espécies, variam em tamanho, formato e textura. Cada folha tem na sua axila dois rebentos, um dos quais na posição central e outro na posição lateral, que é o que dá origem aos ramos frutíferos, enquanto o outro dá origem aos ramos vegetativos (Queiroz, 1965).

Cada ramo frutífero produz seis a oito botões florais, rebentando um em cada nó. Ao fim de vinte e um dias após o seu aparecimento, as flores abrem. As suas pétalas apresentam a cor creme claro, tornando-se rosadas no final do primeiro dia.

O ovário cresce e amadurece até se transformar na cápsula do algodão, que depois de atingir o tamanho natural, demora bastantes dias a amadurecer as suas sementes e fibras, o que só se verifica passados quarenta e cinco a sessenta e cinco dias após a abertura das flores.

Quando a maturação está completa, a cápsula estala e abre-se pelos septos de cada lóculo sob a pressão das fibras, que graças à sua elasticidade se soltam, ficando salientes e constituindo os flocos brancos e felpudos tão característicos das plantações de algodão.

### **3.3 - FISILOGIA DA PLANTA E OS FACTORES QUE INFLUENCIAM O SEU DESENVOLVIMENTO**

O cultivo do algodoeiro é feito em regiões tropicais ou subtropicais, onde a planta recebe um máximo de calor, insolação e humidade, em conjunto com solos férteis e profundos que permitem a expansão das suas raízes.

Embora o algodoeiro possa conservar-se no mesmo terreno durante anos, ainda que a sua vitalidade seja precária e a produção da fibra se apresente insignificante, chegou-se à conclusão que a sua cultura é mais compensadora e racional, quando é cultivado e colhido anualmente, sendo assim possível alternar a sua cultura com outras (Vasconcelos, 1993).



### **i ) Terreno**

O algodoeiro está adaptado a viver em solos permeáveis, nos quais a humidade se encontra no subsolo, pois possui uma raiz principal, que capta a água e o ar que passa através dos solos arenos-argilosos ou argilo-arenosos. As plantas necessitam de grandes quantidades de água, em virtude não só da sua evaporação através das folhas, mas também devido ao facto de a água entrar em elevada percentagem na constituição dos seus tecidos. Assim recorre-se à rega quando a queda pluviométrica não é suficiente, através de diques e canais que transportam a água por distâncias muito grandes, como é o caso do Egipto, do Sudão e algumas regiões da Índia e da América (Queiroz, 1965).

### **ii ) Humidade**

A água sobe pela planta, transportando os sais minerais que são indispensáveis à constituição dos seus tecidos e à formação das fibras e depois evapora-se pelos estomas das folhas.

O facto de a temperatura ser muito elevada nas regiões onde o algodão é cultivado, faz aumentar a evaporação que se realiza à superfície das folhas. No Egipto, por exemplo, a evaporação de um hectare de algodão, atinge as cem toneladas de água, por dia. A planta vai buscar esta grande quantidade de água às reservas existentes nas camadas profundas do subsolo, daí o seu cultivo em solos arenosos ou argilosos, permeáveis à água. Quando o terreno é permeável, mas a distribuição da chuva é irregular ou insuficiente, recorre-se à irrigação, uma vez que a distribuição da chuva durante o período de crescimento da planta influencia mais a sua quantidade e qualidade do que a porção total da chuva caída durante o ano (Vasconcelos, 1993).

### **ii ) Temperatura**

Para se obter algodão de qualidade, é necessário uma temperatura média de 20° a 30° C, chuvas abundantes durante o crescimento e tempo seco e ensolarado na altura da formação dos frutos e da colheita. Daí dizer-se que o algodão deve ter “os pés molhados e a cabeça ao sol”. A temperatura de germinação não deve ser inferior a 20° C, para que as plantas cresçam satisfatoriamente, devendo depois elevar-se até atingir os 30° C. Contudo o algodoeiro adapta-se bem a condições diferentes de temperatura, desde que as exigências relativamente ao tipo de terreno e humidade estejam satisfeitas (Vasconcelos, 1993).



#### **iv ) Pragas**

Existem dois tipos de pragas que atacam o algodoeiro. Uma consiste nos insectos e a outra consiste nas doenças criptogâmicas, que ocorrem onde a humidade do ar e do solo são elevadas, como resultado de excessiva pluviosidade e de a temperatura do ar ser insuficiente para promover a evaporação de forma satisfatória. Como consequência, as plantas ficam enfraquecidas, tornando-se susceptíveis a fungos e bactérias. Este tipo de situação ocorre em certas zonas da América, Sudão e Egipto. Nos E.U.A os prejuízos causados por fungos e bactérias, são sensivelmente superiores aos das zonas tropicais, pois o clima das zonas temperadas faz com que a planta apresente menos resistência às enfermidades e as descidas bruscas de temperatura causam o enfraquecimento das plantas, que é a principal causa das doenças criptogâmicas, como a Bacteriose dos Algodoeiros e a Sintomatologia.

O combate às doenças criptogâmicas tem-se revelado difícil, pois não se conhece nenhum meio prático e económico capaz de eliminar satisfatoriamente os agentes das diversas doenças. Recorre-se à destruição rápida das plantas após a colheita e à obtenção de variedades resistentes ou imunes aos agentes patológicos, através da alteração genética da variedade por adição de DNA de outras variedades mais resistentes ou imunes aos agentes patológicos (Queiroz, 1965).

### **3.4 - CLASSIFICAÇÃO E PADRONIZAÇÃO**

A classificação tem uma importância primordial na comercialização do algodão. Os países produtores de algodão procuram aperfeiçoar a classificação segundo graus de qualidade, criando padrões nacionais e mantendo serviços encarregados de a fazer. Mas nem sempre foi assim, o que tornava difícil o diálogo entre agricultores e comerciantes (Vasconcelos, 1986).

Somente a partir dos princípios do século XIX é que em Liverpool, se começou a distinguir algumas qualidades de algodão que receberam designações especiais.

Em 1874, as Bolsas Algodoeiras Americanas resolveram criar um único tipo de padrões, para vigorar exclusivamente em toda a América do Norte, o qual não deu os resultados esperados por o seu uso não ter sido determinado por legislação oficial. Esta só entrou em vigor em 1909, quando o Departamento da Agricultura dos Estados Unidos resolveu atender às reclamações dos produtores, mandando estabelecer padrões oficiais, por uma comissão de peritos classificadores, formada pelos mais importantes agricultores, comerciantes, industriais e técnicos do governo (Queiroz, 1965).



Estes padrões consistiam em nove graus ou tipos diferentes, dos quais se tiraram diversas cópias fiéis e autenticadas que foram distribuídas pelas escolas de agricultura, escritórios comerciais, fábricas de fiação e pelos principais estabelecimentos interessados em Inglaterra, Alemanha e México.

Em 1914, efectuou-se uma revisão e aperfeiçoamento destes nove graus. Em 1918 foram criados padrões para os algodões Egípcio-Americanos, para os algodões Sea-Island e para algodões de outros comprimentos de fibra.

Finalmente em 1923, houve uma conferência em Washington com os representantes das Associações Algodoeiras de Liverpool, Manchester, Havre, Berma, Roterdão, Barcelona e Milão com o objectivo de uniformizar os padrões americanos com o dos outros mercados mundiais. A partir desta reunião introduziram-se pequenas modificações nos padrões que se passaram a designar por “Universal Standards” (Lucas, 2000).

As amostras eram retiradas dos fardos a uma profundidade suficiente, para se evitar a presença de poeiras e terra vindas do exterior, que facilmente se agarrariam às fibras e deviam ser compostas por uma quantidade de algodão razoável, formadas por camadas o mais intactas e perfeitas possível, para que o aspecto que o algodão apresentasse no fardo não fosse adulterado e prejudicado e para que as amostras fossem suficientemente representativas.

Para se avaliar melhor a fibra, as amostras eram enroladas e embrulhadas em papel de cor azul-escuro para se destacar a fibra, no qual exteriormente se escreviam a origem, a marca e o número do fardo do algodão.

A classificação era feita por comparação do algodão das amostras com os padrões de fibra da mesma origem e era efectuada por peritos classificadores que possuam elevada sensibilidade nos sentidos da visão, tacto e até nos músculos da mão, aliada a uma longa experiência.

Para além da perícia do classificador, a luz e a humidade relativa do local onde era feita a classificação também influenciavam os resultados. Inicialmente as primeiras salas de classificação eram construídas de modo a que a luz fosse recebida por janelas ou clarabóias no tecto ou no alto das paredes, viradas a norte no Hemisfério Setentrional ou a sul no Meridional, de forma a que as amostras nunca recebessem a luz directa do sol, que lhes confere um aspecto uniformemente branco. Mais tarde, técnicos americanos especializados em classificação, encontraram o conjunto de lâmpadas eléctricas que forneciam luz de intensidade semelhante à luz indirecta do sol, permitindo efectuar o trabalho de classificação de forma correcta e sem cansar a vista (Lucas, 2000).



As salas de classificação tinham as paredes pintadas de cinzento claro e instalações de climatização que mantinham constantes a humidade relativa a 65% e a temperatura do ambiente a 20°C. Habitualmente, possuíam também aparelhagem destinada à renovação do ar e à extracção das poeiras, pois não possuíam ventilação suficiente, devido a não terem janelas abertas para o exterior.

O processo de classificação constava de duas fases distintas: determinação do grau e análise da fibra. A determinação do grau era feita por comparação com os padrões pela simples observação à vista da cor, impurezas e a preparação da camada superficial da amostra. Depois, abria-se a amostra com os dedos analisando-se camada por camada, sem as esfarrapar, para se ter a ideia da uniformidade relativa da cor, das impurezas e da preparação na massa do algodão.

Se o aspecto fosse desigual, de camada para camada ou nas duas partes que constituíam uma amostra dupla; isto é, retirada dos dois lados do fardo; era sinal que o fardo analisado era formado por algodão pouco uniforme ou mesmo misturado e então a sua classificação era a do grau que correspondia à parte da amostra que apresentava pior aspecto.

Depois de atribuído o grau, o classificador fazia o “*pulling*” que permitia analisar outras características da fibra como o comprimento e carácter. Este dependia do grau de sedosidade ou aspereza; das ondulações ou convoluções, frisado, espessura, rigidez, flexibilidade e elasticidade; da particularidade das fibras se prenderem umas às outras, quando eram puxadas sobre si, ou de se soltarem facilmente; da consistência ou resistência, que evita a rotura das fibras, a qual faz aumentar os desperdícios da fiação; da regularidade ou desigualdade no comprimento e do brilho ou lustro.

O “*pulling*” permitia obter um feixe de fibras cujo comprimento era avaliado por comparação, com outro feixe idêntico, extraído do algodão dos padrões de comprimento, ou então por medida directa sobre uma régua graduada de preferência transparente (Queiroz, 1965). Os padrões para o comprimento eram os padrões americanos constituídos por pacotes de fibras, com rótulo, carimbo e assinatura dos Departamentos da Agricultura dos Estados Unidos, embrulhados em papel parafinado e autenticados.

Nos últimos cinquenta anos, foram desenvolvidos instrumentos capazes de determinar algumas propriedades das fibras com um grau aceitável de reproductibilidade. A automatização dos ensaios às fibras de algodão começou em 1965, resultado da colaboração entre o Departamento de Agricultura dos Estados Unidos e o fabricante americano de instrumentos *Motion Control* de Dallas. Esta colaboração teve como resultado a primeira linha de Instrumentos de Alto Volume



(HVI), que seria seguida por um segundo sistema similar desenvolvido em 1984 pelo Spinlab (Furter et al., 1990). Desde 1969, o sistema *Motion Control* HVI tem sido o sistema padrão usado pelos classificadores Americanos. Desde então as unidades de HVI têm sido continuamente melhoradas, tanto em *software* como em *hardware* (Vasconcelos, 1993).

Os aparelhos de alto volume (HVI) permitem realizar ensaios que determinam a Finura, Comprimento de fibra, Uniformidade, Resistência, Cor e Percentagem de Impurezas. Uma das suas principais vantagens é o conhecimento das propriedades físicas da fibra de algodão, a obtenção de resultados precisos e reproduzíveis, dados imediatamente disponíveis e a utilização dos resultados para prever a qualidade do fio (Parker, 1982). O sucesso na operação de fiação é cada vez mais da responsabilidade do Laboratório Têxtil que fornece a base para a tomada de decisão sobre o uso efectivo dos materiais, por isso o controlo de qualidade contínuo e a selecção de matéria-prima são indispensáveis (Muller et al., 1991).

Os produtores preocupam-se com a qualidade da fibra quando as entidades classificadoras os informam que o valor comercial de determinado algodão baixou, porque as suas propriedades apresentam valores que se encontram fora dos limites aceitáveis ou estabelecidos. Os industriais preocupam-se com esta questão quando surgem problemas na fiação ou defeitos significativos no fio ou no artigo acabado devido à baixa qualidade da matéria-prima (Bradow et al., 2000). Fardos de baixa qualidade, que não podem ser convenientemente processados são devolvidos aos produtores que não têm meios disponíveis para cobrir os custos do algodão rejeitado.

### **3.5 - ESTRUTURA MORFOLÓGICA DA FIBRA**

As fibras de algodão são células do tegumento da camada exterior das sementes do algodoeiro, que se diferenciam no sentido do comprimento, hipertrofiando-se e prolongando-se até atingirem 15 a 40 milímetros, conforme a variedade em questão (Queiroz, 1965).

Logo que os óvulos da flor são fecundados pelos gâmetas masculinos transformam-se em pequenas sementes que se formam muito depressa, atingindo o tamanho natural em aproximadamente vinte e cinco dias. Se durante este período ocorrerem quaisquer deficiências nas condições de clima ou de alimentação da planta, o crescimento da fibra pode ser prejudicado e esta ficar com um comprimento inferior ao normal, característico da variedade em questão.



Na primeira fase de crescimento, a fibra apresenta-se como um cilindro oco cuja camada interna é formada por celulose e recebe a designação de “camada primária”, enquanto a camada externa é formada por uma cutícula recoberta de ceras.

Na segunda fase de crescimento que corresponde à maturação da cápsula e demora entre vinte a trinta e cinco dias, forma-se interiormente uma segunda camada que tem a designação de “camada secundária” que engrossa rapidamente por deposição de celulose em camadas concêntricas diariamente até quatro ou cinco dias antes do rebentamento da cápsula. A estrutura da camada secundária nunca enche a célula completamente, pelo que existe sempre algum espaço no centro da fibra que é designado por lúmen.

O processo de crescimento termina com a abertura da cápsula que estala. As fibras que até então eram cilíndricas, cheias de protoplasma, aderentes e dobradas, sem resistência nem elasticidade, principiam a secar pela evaporação da grande quantidade de água que contém, passando o seu interior a ficar vazio, o que origina que as paredes se enrijeçam, se contraíam e encostem entre si.

Com a secagem do protoplasma as pequenas perfurações ou cavidades oblíquas que existem nas paredes das fibras tapam-se e dão origem a um enrugamento nesses pontos o que causa curvaturas de 180° relativamente ao eixo da fibra, num número que depende principalmente da rigidez da parede, do tamanho do lúmen e diâmetro da fibra. Estas curvaturas são designadas por convoluções. Por este facto as fibras que prematuramente secaram ou morreram, apresentam-se lisas, sem ondulações ou sinuosidades e com falhas de resistência em virtude de terem secado antes que se completasse a formação das paredes (Lucas, 2000).

Os trabalhos de Balls na década de vinte, de Fan e Kerr nas de trinta e quarenta, de Flint na de cinquenta contribuíram de forma decisiva para o conhecimento da estrutura da fibra de algodão (Hamby, 1965). Estes autores efectuaram análises químicas com certos reagentes e indicadores com afinidade para a celulose, para o conteúdo em pectinas ou para os complexos de cinzas e visualizaram a estrutura da fibra de algodão. Demonstraram que a arquitectura natural da fibra de algodão é a de cilindros concêntricos encaixados numa membrana ou camada primária que contém substâncias não celulósicas.

A primeira subcamada da camada secundária apresenta uma organização diferente das restantes camadas e também da camada primária. A última subcamada da camada secundária constitui a fronteira para o lúmen ou canal central.



### 3.5.1 - Corte Longitudinal

Longitudinalmente, a fibra madura apresenta três partes (Hamby, 1965):

**i ) Base:** trata-se de uma porção frágil e curta, com a forma cônica que durante o crescimento permanece entalada entre as células da epiderme. Geralmente esta base é arrancada na colheita pelo que raramente é encontrada nas fibras de algodão comercial.

**ii ) Corpo:** é a porção principal da fibra de algodão, que pode medir três quartos a quinze dezasseisavos do seu comprimento e cujo diâmetro se mantém com uma parede dura e um lúmen estreito. Só na extremidade é que o diâmetro diminui e o lúmen deixa de existir.

**iii ) Extremidade:** trata-se de uma porção estreita e dura no final da fibra que pode medir até um quarto do comprimento total da fibra mas que é geralmente menor. Não contém lúmen, o seu diâmetro é consideravelmente menor que o do corpo da fibra e apresenta poucas convoluções.

### 3.5.2 - Corte Transversal

Numa fibra madura distinguem-se seis partes (Hamby, 1965):

**i ) Cutícula:** é uma camada exterior, muito fina, não celulósica, constituída por ceras, mas também pode conter pectinas e material proteico e funciona como uma capa macia e protectora, impermeável à água, que protege o resto da fibra da oxidação quando exposta aos raios ultravioleta.

**ii ) Camada Primária:** é a camada original da célula. É fina e constituída principalmente por celulose, mas contém pectinas, proteínas e ceras, sendo impregnada e coberta pelos materiais que compõem a cutícula.

**iii ) Capa da camada secundária:** é uma espécie de pele da camada secundária e difere de alguma forma da estrutura da camada primária e do resto da camada secundária.

**iv ) Camada secundária:** consiste em subcamadas concêntricas de celulose que constituem a parte principal da fibra de algodão. Devido à elevada cristalinidade, a camada secundária é responsável pela resistência mecânica da fibra.

**v ) Capa do lúmen:** camada que aparenta ser mais resistente a certos reagentes que a camada secundária.



vi ) **Lúmen:** é o espaço vazio dentro da fibra que depois do colapso forma um canal que varia em tamanho e forma e ocupa diferentes posições ao longo da mesma fibra. Frequentemente o lúmen contém matéria sólida como substâncias proteicas, sais minerais e matérias corantes que constituem restos do protoplasma seco.

### 3.6 - PROPRIEDADES FÍSICAS

A Física desempenha um papel importante na Tecnologia Têxtil. A Física Têxtil estuda a estrutura e propriedades físicas das fibras. A matéria é composta por átomos ligados por forças de ligação de intensidade variável. É o arranjo destes átomos e as suas forças de ligação que determinam as propriedades físicas dos materiais (Lucas, 2000).

Nos Estados Unidos, todos os fardos são analisados pelos gabinetes de classificação do Departamento dos Estados Unidos e os parâmetros de qualidade da fibra de algodão são definidos pelas propriedades físicas mensuradas, que são habitualmente o Comprimento, o Índice de Uniformidade, a Resistência, o Índice Micronaire, a Cor, traduzida pela Reflectância (Rd) e Grau de Amarelo (+b) e a Percentagem de Impurezas, todas quantificadas pelos sistemas HVI (Bradow et al., 2000).

A qualidade ideal para a fibra de algodão, apresentará as especificações que sabiamente os produtores preconizavam:” *as white as snow, as long as wool, as strong as steel and as fine as silk* “ (Bradow et al., 2000), ou seja “tão branca como a neve, comprida como a lã, resistente como e aço e fina como a seda”.

#### 3.6.1 – Comprimento

Este parâmetro é dos mais importantes na caracterização de um determinado algodão, devido ao facto de que na comercialização da rama, bem como no seu processamento, ser um dos responsáveis pelo custo da matéria-prima e pelo bom desempenho do processo (Vasconcelos, 1993).

Devido à variabilidade inerente à fibra de algodão não existe um valor absoluto para o comprimento da fibra de determinado genótipo (Bradow et al., 2000). Na mesma semente, o comprimento das



fibras varia significativamente; as fibras mais compridas encontram-se na parte mais larga da semente e as mais curtas na sua parte mais alongada.

No tipo *Upland*, o coeficiente de variação do Comprimento de fibra pode atingir valores da ordem dos 40%, coeficientes estes que variam de amostra para amostra.

As variações no comprimento de fibra atribuídas ao genótipo e à posição da fibra na semente são condicionadas por factores micro e macroambientais. Mudanças ambientais ocorridas durante o período de floração da planta, podem limitar o início da formação das fibras ou atrasar o seu hiperatrofiamento e alongamento.

Os produtores e compradores usavam o termo “*Staple Length*” (comprimento do género principal), quando se referiam ao comprimento, muito antes do aparecimento de métodos de ensaio para quantificação desta propriedade.

Em termos históricos, o Comprimento de fibra começou por ser medido usando o Diagrama de Baer ou o método de Suter-Webb. Ambos baseavam-se em dispor as fibras de determinada amostra, segundo o seu comprimento ou peso (Bradow et al., 2000). As fibras eram penteadas, paralelizadas e separadas em grupos de 1/8 de polegada de amplitude.

No método de Suter-Webb, as fibras de cada grupo eram pesadas. A distribuição comprimento/peso era usada para calcular diversas propriedades relacionadas com o Comprimento, incluindo a média do comprimento das fibras e o valor do 3º Quartil.

A construção do Diagrama de Baer era feita manualmente, pelo que o método depressa se tornou proibitivo em termos de mão-de-obra e tempo, principalmente nos gabinetes de classificação e unidades industriais.

Como o Comprimento de fibra está directamente relacionado com a produtividade da fição, massa linear, resistência e uniformidade do fio, foram desenvolvidos métodos instrumentais para medição do Comprimento de fibra, rápidos e reproduzíveis.

No Fibrógrafo, as fibras são presas aleatoriamente a pentes e o conjunto é “varrido” fotoelectricamente desde a base até ao topo. A quantidade de luz que passa através do diagrama é a medida do número de fibras que estão a determinada distância dos pentes. Os dados são registados num gráfico e o “*Span Length*” mostra a distância marcada por uma dada percentagem de fibras no Fibrógrafo.



Um dos valores típicos do gráfico é a determinação do *Span Length* a 2,5% que é o valor do comprimento calculado para esta percentagem de fibras e corresponde às fibras mais compridas. É geralmente semelhante ao do comprimento da fibra obtido pelo “*pulling*” +/- 1 milímetro e é utilizado para afinar a distância entre os cilindros de estiragem, enquanto o *Span Length* a 50% é uma medida do potencial do processo de fição e da qualidade do fio (Vasconcelos, 1993).

Comprimento ( cm )	Classificação
< 0,99	<b>Curta</b>
0,99 – 1,10	<b>Média</b>
1,11 – 1,26	<b>Longa</b>
> 1,26	<b>Extra longa</b>

**Tabela 2** – Classificação das fibras quanto ao Comprimento (Vasconcelos, 1993).

Outra das propriedades importantes para o fiandeiro é o índice de uniformidade do comprimento de fibra, dado este traduzir a forma de distribuição do comprimento de fibra, pois uma grande variedade no comprimento tende a aumentar os desperdícios da fição para além de provocar perturbações no processo (Schenek et al., 1990). Este é normalmente expresso como sendo a razão entre o comprimento médio e o comprimento médio da metade superior. Os algodões classificam-se da seguinte forma:

Índice de Uniformidade ( % )	Classificação
< 77	<b>Muito baixo</b>
77 – 79	<b>Baixo</b>
80 – 82	<b>Médio</b>
83 – 85	<b>Elevado</b>
> 85	<b>Muito elevado</b>

**Tabela 3** – Classificação das fibras quanto ao índice de uniformidade (Vasconcelos, 1993).



Desde 1980, o Departamento de Agricultura dos Estados Unidos, confia quase totalmente nos Instrumentos HVI para a medição do Comprimento de fibra e outras propriedades físicas.

Amostras de algodão padrão são utilizadas para que os testes que se realizam nas unidades industriais e outras possam ser validados e aceites (Bradow et al., 2000).

Todos os métodos de análise mencionados requerem um mínimo de 5 gramas de massa de fibras, sendo utilizados para uma classificação rápida de grandes volumes de fibras. Contudo quando se pretende analisar pequenas amostras de fibras, por exemplo, fibras de uma semente ou de uma cápsula, utiliza-se o sistema AFIS A2 – *Advanced Fiber Information System*, da *Zellweger Uster*, que mede parâmetros como a média do Comprimento de fibra por massa e por número, constrói histogramas do Comprimento de fibra, determina os valores correspondentes ao 3º Quartil e a Percentagem de fibras curtas por massa e por número.

Apesar da percentagem de fibras curtas (proporção de fibras curtas que é especificada como uma percentagem do peso das fibras mais curtas que 12,7 milímetros) não ser incluído na classificação oficial do USDA (Departamento de Agricultura dos Estados Unidos), o reconhecimento da sua importância como uma propriedade da fibra, semelhante à importância de propriedades como a Finura, Resistência e Comprimento, tem vindo a aumentar. Um aumento na percentagem de fibras curtas pode causar um aumento na percentagem de desperdícios da fiação, uma menor resistência do fio, assim como um aumento da sua irregularidade (Lucas, 2000).

Enquanto o Comprimento é praticamente determinado pelo factor genético, a percentagem de fibras curtas para além de ser condicionada pelo factor genético depende também das condições ambientais, da colheita e do descaroçamento.

É essencial que genicistas, fisiologistas, produtores e construtores de aparelhos de medição e controlo percebam os conceitos e ultrapassem as limitações dos métodos de quantificação do Comprimento de fibra e da percentagem de fibras curtas, para que a forte componente do factor genético nesta propriedade possa ser separada dos factores ambientais, introduzidos por deficiências nas condições de temperatura, água, luz e constituição química do solo. Faz-se referência à influência destes factores ambientais no Comprimento de fibra (Lucas, 2000):

**i ) Temperatura:** Comprimentos de fibra máximos foram atingidos quando a temperatura nocturna rondava os 19° C a 20° C, dependendo da variedade.



O Comprimento de fibra foi negativamente correlacionado com diferenças entre a temperatura máxima e a temperatura mínima. Experiências mostraram que temperaturas noturnas de 15°C, causaram quatro a cinco dias de atraso no hiperatrofiamento e posterior alongamento da fibra, relativamente a temperaturas de 25°C. Temperaturas altas podem promover a diminuição ou até ausência de cápsulas pequenas bem como a diminuição da percentagem de fibras curtas.

**ii) Água:** Grimes e Yamada concluíram que o comprimento de fibra não é afectado a não ser que a carência em água seja suficiente para baixar a produtividade da colheita para 700 quilogramas por hectare. A ocorrência de deficiências de humidade no início do período de floração, não altera o Comprimento de fibra, contudo se a secura ocorrer mais tarde, o comprimento da fibra diminui. Carências durante o período de alongamento reduzem o comprimento da fibra, aparentemente devido ao processo mecânico e fisiológico de expansão da célula.

**iii ) Luz:** Mudanças de luminosidade no ambiente de crescimento alteram a configuração da planta. A quantidade de luz é importante para a realização da fotossíntese pela planta. Experiências em que a quantidade de luz era reduzida aumentaram o comprimento das fibras. Sombra durante os primeiros 7 dias de período de floração, resultaram em aumentos de 2% no *Span Length* de 2,5% para certos genótipos.

**iv ) Composição do solo:** Investigações sobre a nutrição em minerais do algodoeiro e a composição química dos solos, enfatizaram a eficácia da frutificação e a produtividade da colheita. Estas investigações estudaram o efeito da presença de potássio e azoto nas propriedades da fibra, incluindo o *Span Length* e verificaram-se efeitos contraditórios que dependem da acção interactiva do genótipo, clima e composição do solo.

Crê-se que o Comprimento de fibra é determinado em grande parte pelo genótipo, mas flutuações no ambiente de crescimento, resultantes da variabilidade das condições climatéricas ao longo das estações ou induzidas pelos métodos utilizados na cultura, colheita e preparação da fibra, moldam a amplitude e média do comprimento de fibra na população de uma amostra, fardo ou colheita.

### 3.6.2 - Propriedades Dinamométricas

As propriedades dinamométricas das fibras têxteis – a resposta a forças e deformações – são provavelmente as mais importantes uma vez que contribuem tanto para o comportamento das fibras durante o processamento, como para as características do produto final (Morton et al., 1975).



Consideram-se como propriedades dinamométricas, a resistência, a tenacidade e o alongamento (Vasconcelos, 1993). A resistência das fibras individuais é considerada o parâmetro mais importante na determinação da resistência de um fio (Bradow et al., 2000).

### 3.6.2.1 - RESISTÊNCIA E TENACIDADE

Em termos históricos, a resistência da fibra podia ser quantificada por dois tipos de aparelho, o *Pressley* e o *Stelometer*. Em ambos os aparelhos um feixe de fibras paralelas é colocada nas maxilas do aparelho, depois de penteadas, separadas por um espaçador de 1/8 de polegada ou então utilizando-se um espaçamento igual a zero e submetendo-se as fibras a uma força de tracção crescente até provocar a sua ruptura (Vasconcelos, 1993).

Geralmente, para a distância de zero entre as amarras, o que ainda é muito comum, os resultados são expressos em termos de unidades *Pressley* (mil libras por polegada ao quadrado), obtendo-se a seguinte escala de valores para as fibras de algodão.

Escala de valores Pressley	Classificação
> 93	Excelente
92 – 87	Muito forte
86 – 81	Forte
80 – 75	Médio
74 – 70	Fraco
< 70	Muito fraco

**Tabela 4** – Classificação do algodão quanto à Resistência (Klein, 1980).

De qualquer modo, a tendência é de se utilizar o espaçador (distância entre amarras de 1/8”), pois o valor assim obtido é mais fácil de se correlacionar com a resistência do fio (Bradow et al., 2000). Geralmente este teste mede a tenacidade (resistência/massa por unidade de comprimento), é efectuado no *Stelometer* e expresso nas unidades *gf/tex* ou *cN/tex*. Um algodão resistente pode atingir valores de Tenacidade de 25 *cN/tex*.

A classificação do algodão, utilizando o *Stelometer* com espaçador é a seguinte:



Tenacidade ( cN/tex )	Classificação
< 17	Muito fraco
18 – 21	Fraco
22 – 25	Médio
26 – 29	Resistente
> 30	Muito resistente

**Tabela 5** – Classificação das fibras quanto à Tenacidade (Vasconcelos, 1993).

Os valores para o algodão são os mais elevados de entre as fibras naturais vegetais mas são inferiores aos da maioria das fibras regeneradas e das naturais animais. Parâmetros como a orientação molecular e o ângulo das espirais estão significativamente correlacionados com a Tenacidade (Hamby, 1965).

A percentagem de humidade tem um efeito considerável na resistência e o seu efeito depende do estado do algodão, do seu genótipo e maturidade. A resistência das fibras de algodão, individuais ou em feixe, aumenta com a humidade relativa até 60%. A partir deste valor não regista mais nenhum aumento. Todavia, se as fibras estão molhadas um novo aumento pode acontecer e pode ir até 20%.

Crê-se que as diferenças de resistência entre diferentes tipos de algodão podem ser atribuídas em grande parte, se não totalmente, a diferenças na estrutura e na morfologia destes, tais como o comprimento e a orientação das cadeias moleculares da celulose e os ângulos das convoluções.

A Rigidez pode ser definida como a resistência à deformação e para os têxteis é habitualmente definida como a razão entre a tracção de ruptura e a força de tracção.

De acordo com Betrabet, a rigidez representa o efeito combinado do peso molecular, da rigidez natural das cadeias de celulose, o grau de cristalinidade formado pelas pontes de hidrogénio e as forças de Van der Waals e a orientação das cadeias de celulose relativamente ao eixo da fibra (Hamby, 1965). O coeficiente de correlação entre o ângulo das espirais e a rigidez é de 0.45 a 0.65, o que demonstra que a orientação das cadeias são um factor importante na rigidez.



### 3.6.2.2 - ALONGAMENTO

O alongamento representa a extensão do material quando este é traccionado até à ruptura. Apesar de poder ser expresso em unidades absolutas, é habitualmente traduzido por uma percentagem do comprimento inicial (Hamby, 1965).

O algodão é o que apresenta os valores mais altos entre as fibras naturais vegetais e é em média o dobro da fibra que se lhe segue, o ramie. Apenas as fibras naturais animais, a lã e a seda ultrapassam o algodão em termos de alongamento. O valor médio é de 7%, podendo variar entre 4 e 13%, dependendo da variedade em questão. Foi demonstrado que o alongamento aumenta à medida que a orientação axial diminui. A classificação dos algodões quanto ao alongamento é a seguinte:

Alongamento ( % )	Classificação
< 5	Muito Baixo
5,0 – 5,8	Baixo
5,9 – 6,7	Médio
6,8 – 7,6	Elevado
> 7,6	Muito elevado

**Tabela 6** – Classificação do algodão quanto ao Alongamento (Parker, 1982).

### 3.6.3 - Finura e Maturidade

Das propriedades definidas pelo Departamento de Agricultura dos Estados Unidos, a maturidade da fibra é provavelmente a mais ambígua (Bradow et al., 2000). O termo maturidade da fibra usado no mercado do algodão e seu processamento, não é apenas uma estimativa do tempo decorrido entre o início da floração e a colheita da fibra, mas sim uma combinação de factores genéticos e de crescimento, assim como a finura da fibra e a razão entre a espessura da camada primária e a da camada secundária.

A finura da fibra (densidade linear da fibra) é expressa em função de uma massa por unidade de comprimento. Esta propriedade vai determinar quantas fibras se podem encontrar na secção transversal do fio (Vasconcelos, 1993). Um aumento do número de fibras na secção transversal, para além de aumentar a resistência faz com que estas se distribuam melhor no fio. A finura não só



determina o limite de fiação, como também influencia a resistência e uniformidade do fio e a resistência ao enrugamento, textura, brilho e caír do produto final (Muller et al., 1991).

Para um dado tipo de algodão, o diâmetro original da fibra é relativamente constante. Contudo, as paredes da célula variam consideravelmente em grau de espessura, o que provoca extrema variabilidade tanto no tamanho como na forma da secção de um corte transversal da fibra.

A maturidade corresponde ao grau de desenvolvimento das camadas de células da fibra de algodão e pode ser medido pela proporção da espessura da parede relativamente à largura da fibra. Com tantas fibras numa só semente, a competirem pelos nutrientes, é compreensível que não desenvolvam todas a mesma espessura para as suas paredes.

A forma geométrica habitual para as fibras maduras é elíptica a circular, geralmente em forma de feijão e para as fibras mais finas, é plana, rectangular, com os cantos arredondados. As fibras imaturas apresentam geralmente a forma de “U” devido à tendência das fibras de paredes finas enrolarem-se em torno de si mesmas na secagem, depois do colapso ou rebentamento da cápsula (Hamby, 1965).

Para além disso e como já foi referido, as fibras não começam todas o seu desenvolvimento no mesmo dia, e as condições de crescimento, tempo de colheita, ataques de doenças podem impedir o algodão de atingir a sua maturidade máxima, obtendo-se desta forma uma fibra imatura. Cerca de um quinto do número total de fibras numa amostra de algodão, apresentar-se-á subdesenvolvido ou imaturo (Hamby, 1965).

Estas fibras imaturas (por vezes designadas por “fibras mortas”) mesmo com um comprimento normal, mostram pouco ou nenhum desenvolvimento da camada secundária e depois da abertura da cápsula, estas aparecem com a forma de fitas, com a secção transversal mais achatada (Queiroz, 1965). Por vezes, numa amostra de algodão, verifica-se que paredes finas e lúmen grande, redondo e vazio pertencem a fibras em que não ocorreu o colapso. Este facto está geralmente associado a estações de chuvas abundantes e consequentes condições de humidade evitaram a secagem rápida e o colapso ou então a cápsulas apodrecidas.

Peirce (Morton et al., 1975) definiu a maturidade como a razão entre a secção transversal (A) e a área do círculo com o mesmo perímetro (A0). Matematicamente, ter-se-á:



$$\text{Maturidade} = \frac{A}{A0}$$

**Equação 1** – Definição de Maturidade (Vasconcelos, 1993).

Sendo a maturidade expressa em termos de percentagem.

Schneck (Klein) sugere que a fibra deve ser considerada madura, quando a parede celulósica representa 50% a 80% da sua secção transversal, imatura quando representa 30% a 45% e morta quando o valor for inferior a 25%. O aparecimento de 5% de fibras imaturas num fardo, pode contribuir para a diminuição da resistência do fio, um maior número zonas finas, aumento no número de fibras curtas, quebras na fição, excessos de desperdícios e defeitos no tingimento como pontos brancos e barrado, originando tingimentos irregulares (Vasconcelos, 1993).

As fibras imaturas apresentam geralmente uma espessura da ordem dos 2 milimícron.

De qualquer forma, é necessário ter em conta que as cápsulas de algodão não contêm só fibras maduras. Mesmo que o seu crescimento tenha ocorrido nas melhores condições, contêm pelo menos 0.5 % de fibras imaturas.

O grau de maturidade das fibras de algodão tem importância no comportamento da matéria-prima durante o processamento. A afinidade das fibras aos corantes é directamente proporcional ao grau de maturidade das fibras.

O quadro seguinte refere-se à maturidade, expressa em percentagem:

Maturidade ( % )	Classificação
< 0,7	Muito imatura
0,7 – 0,8	Imatura
0,8 – 1,0	Madura
> 1,0	Muito madura

**Tabela 7** – Classificação do algodão quanto à Maturidade (Vasconcelos, 1993).



O índice micronaire, dentro de uma variedade de algodão é geralmente a medida da sua maturidade (Balasubramanian, 1989). Mas quando se trata de diferentes variedades de algodões, o índice micronaire é uma função comum, tanto da maturidade como da finura intrínseca ou perímetro do algodão.

Na prática, o efeito da mudança de índice micronaire, pode ser devido a dois factores: se há mudança na maturidade da fibra (factor ambiente) ou então se há mudança na finura básica do algodão (factor genético) (Vasconcelos, 1993). A cutícula e a camada primária são responsáveis por cerca de 2,4 % da espessura total da fibra, correspondendo os restantes 98 % à camada secundária que engrossa rapidamente com a fotossíntese e a deposição de subcamadas de celulose, durante o período de maturação da fibra. Por conseguinte, qualquer factor ambiental que afecte a actividade fotossintética irá condicionar o espessamento e consequente maturação fisiológica da fibra.

O valor obtido para o índice micronaire é muito utilizado na indústria como uma medida da finura da matéria-prima, talvez por ser fácil determinação.

Índice Micronaire	Finura
< 3,5	<b>Muito fina</b>
3,5 – 3,9	<b>Fina</b>
4,0 – 4,9	<b>Média</b>
5,0 – 5,9	<b>Grossa</b>
> 6,0	<b>Muito grossa</b>

**Tabela 8** – Classificação do algodão quanto ao Índice Micronaire (Ferreira, 1960).

### 3.6.4 – Cor

O algodão nunca é verdadeiramente branco. Todas as amostras de algodão apresentam uma coloração creme claro. Alguns dos algodões de fibra longa do género *Gossypium Barbadense* como os do Egipto são mais cremes que os algodões do género *Hirsutum* como os Americanos *Upland* e os de Sea Island (Hamby, 1965).

A natureza do pigmento responsável pela cor não é conhecida. A fibra de algodão aumenta a coloração após um período de armazenagem de dois a cinco anos.



Alguns algodões apresentam colorações fortes. A sua gama varia de cor caramelo escuro à caqui e ao bege. Existem algodões verdes cuja cor de azeitona desbota após exposição à luz. A cor creme natural dos diversos graus de algodão comercial pode ser destruída por condições climáticas severas ou por contaminação por poeiras, sujidade e insectos fazendo com que a cor se torna acinzentada e perca brilho. Uma descoloração branco azulada que não desaparece no branqueamento é usualmente atribuída ao desenvolvimento de fungos durante uma exposição excessiva ao clima no campo antes da colheita.

Hoje em dia a cor é objectivamente medida num instrumento designado por Colorímetro e expressa em termos de grau de cinzento ou amarelo (Vasconcelos, 1993). A amostra é iluminada por duas lâmpadas de igual energia e espectro específico a 45 graus. A luz reflectida pela amostra é conduzida através de dois filtros com uma cortina de absorção espectral e é determinada a energia residual recebida nas duas fotocélulas (Muller et al., 1991). O volume de energia nas fotocélulas é expresso nos valores *tristimulus* do sistema CIE – Comissão Internationale del' Eclairage – e depois convertidos.

Dois parâmetros estão relacionados com a medição da cor, a reflectância (Rd) que indica o grau de cinzento da amostra e o Grau de Amarelo (+b) em que a dominante amarelada da amostra é medida utilizando um filtro amarelo. O valor da reflectância pode variar entre 0 % (preto) e 100 % para o branco absoluto, portanto quanto maior a reflectância mais branca é a amostra. Quanto maior o grau de amarelo, mais amarelada é a amostra.

Das três componentes do grau do algodão – cor, percentagem de impurezas e preparação – a cor é a que mais se relaciona com factores ambientais.

Na maior parte dos casos quando o grau de cinzento é baixo, o branqueamento pode reduzir ou até eliminar a cor existente no algodão, mas também pode fazer com que a resistência deste diminua.

### **3.7 - CARACTERÍSTICAS QUÍMICAS**

O quadro refere-se às características químicas de fibras maduras de algodão de diferentes origens, em matéria seca (Hamby, 1965):



Constituinte	Empire	Wilds	Stoneville	Desconhec. Madura	Desconhec. Imatura	Desconhec. (Mebane)
Celulose (%)	95,30	93,42	95,58	96,41	92,44	94,93
Teor de Açúcares (%)	0,07	0,25	0,18	0,17	0,13	0,11
Açúcares redutores (%)	0,01	0,17	0,04	0,02	0,03	0,02
Azoto (%)	0,17	0,26	0,18	0,16	0,32	0,19
Ceras (%)	0,73	1,07	0,69	0,45	1,14	0,57
Cinzas (%)	0,86	1,32	1,00	0,79	1,32	1,16
Alcalinidade das cinzas	13,97	20,52	14,81	11,75	21,75	16,46
PH do extracto aquoso	7,16	6,23	7,50	7,58	5,84	6,46
Extracção com água (%)	2,86	3,51	2,33	2,00	4,62	3,02

**Tabela 9** – Características de algodões de diferentes origens, em matéria seca (Vasconcelos, 1993).

### 3.7.1 – Celulose

Sabe-se que a camada secundária é praticamente constituída por celulose, que é um polímero natural derivado da condensação da **B** – glucose ligada nas posições carbono um e carbono quatro. Em cada unidade de glucose podem ser observados três grupos hidróxilo, um primário e dois secundários. A polaridade dos grupos é responsável pelas fortes forças intramoleculares – pontes de hidrogénio – entre as várias moléculas (Aguiar, 1997).

As cadeias são compostas por cerca de  $10^4$  unidades de glucose. A cadeia completa mede cerca de 5 milimícron de comprimento por  $8 \times 10^{-4}$  milimícron de largura (Morton et al., 1975).

As outras substâncias pertencem à camada primária e ao resíduo protoplásmico do lúmen. A camada primária também contém bastante celulose, ceras, substâncias pectinas, algumas das cinzas e parte do material com azoto (tabela 10).



Constituinte	Camada Primária	Fibra
Celulose	54	94
Proteínas	14	1.3
Pectinas	9	1.2
Ceras	8	0.6
Cinzas	3	1.2
Outros	12	1.7

**Tabela 10** – Composição da camada primária e da fibra, em percentagem (Lucas, 2000).

Os pigmentos, o resto das proteínas, das cinzas, açúcares, ácidos orgânicos são encontrados nos resíduos do lúmen. As ceras distribuem-se pela camada primária de uma forma que tornam as fibras de algodão hidrófobas e contribuem para o aumento da sua resistência à abrasão (Hamby, 1965).

### 3.7.2 – Pectinas

De entre os constituintes não celulósicos da fibra de algodão, as pectinas ocupam o primeiro lugar. O conteúdo em pectinas numa fibra madura é de cerca de 0,6 a 1,2%, dependendo da exactidão do método de determinação (Hamby, 1965).

Muitas das pectinas estão localizadas na camada primária conforme o indicado por coloração, mas a análise de amostras da camada primária removida de fibras maduras indica que esta não é a única responsável pelo conteúdo das pectinas da fibra.

Quimicamente, as pectinas são cadeias de carboidratos de alto peso molecular, semelhantes à celulose.

As pectinas são removidas do algodão nas operações de tratamento prévio. A sua remoção não altera a fluidez em hidróxido de cupramónio nem a sua resistência.



### **3.7.3 - Substâncias proteicas**

O material proteico no algodão é constituído pelos restos do protoplasma, deixado no lúmen após a morte da célula quando a cápsula estala.

Tripp encontrou azoto na parede primária retirada de algodão madura, numa percentagem de 2% que calcula ser aproximadamente 14% da composição proteica na camada exterior da fibra (Hamby, 1965). A operação de fervura reduz o conteúdo em azoto para cerca de um décimo do valor inicial.

Infelizmente, pouco é conhecido acerca da composição e propriedades das proteínas da fibra de algodão ou acerca da natureza dos outros constituintes com azoto na fibra.

### **3.7.4 - Ceras**

É o material extraído das fibras de algodão com solventes orgânicos, como o clorofórmio, benzeno, tetracloreto de carbono, éter de petróleo entre outros. As ceras extraídas do algodão verde são duras, castanhas e em muitos aspectos semelhantes à cera fabricada pelas abelhas.

No algodão caroço, as ceras actuam como uma capa protectora repelente da água, fazendo com que este possa flutuar na água durante dias, enquanto o algodão sem ceras retiradas por aquecimento com hidróxido de sódio diluído ou por extracção com solventes orgânicos, afundar-se-á em poucos minutos (Hamby, 1965).

Do ponto de vista do processamento, as ceras são o constituinte mais importante, depois da celulose, pois a sua presença é necessária para a fiação, uma vez que lubrificam a fibra. As ceras naturais do algodão diminuem a tendência das fibras para aderirem umas às outras, reduzem as fricções entre as fibras e baixam os valores de resistência de fios e tecidos. Aumentos de cerca de 25% na resistência ocorrem em fios ou cordas feitos de algodão ao qual foram extraídas as ceras com benzeno ou outros solventes.

As ceras não são totalmente removidas no processamento têxtil, uma vez que investigações efectuadas por Kettering e Kraemer, em tecidos fervidos e branqueados, quer pelo branqueamento com peróxido de hidrogénio ou pelo branqueamento com hipoclorito de sódio, demonstraram percentagens de 0,3 a 0,5% de Ceras nesses tecidos (Lucas, 2000). A fibra madura apresenta normalmente 0,6% de ceras, numa gama de valores compreendidos entre 0,4% e 1,3%. As ceras localizam-se praticamente todas na camada primária, contudo não se sabe se constituem uma capa



que envolve a camada primária, ou se estão combinadas quimicamente com outros constituintes como as pectinas, celulose ou proteínas. Pequenas quantidades de ácidos gordos como o palmítico, estérico e oleico são também encontrados nas ceras (Hamby, 1965).

### 3.7.5 - Cinzas

O algodão retirado directamente do descaroçador apresenta uma percentagem de cinzas entre 2 a 3%, provavelmente devido a fragmentos da cápsula, bem como a poeiras do solo que se aderem às fibras (Hamby, 1965).

Análises às cinzas demonstram que os seus principais constituintes são o magnésio, cálcio, carbonato de potássio, fosfatos, sulfatos ou cloretos. Estes carbonatos são resíduos de metais originalmente presentes ou de ácidos orgânicos.

As operações de fervura e branqueamento reduzem as cinzas para menos de 0,1%.

As cinzas são fortemente alcalinas, uma propriedade típica das cinzas das plantas. Cerca de 85% do teor das cinzas podem ser removidas da fibra por extracção com água, mas a maior parte dos sais de cálcio, ferro e alumínio permanece na fibra (Lucas, 2000).

### 3.7.6 – Açúcares

Os açúcares contribuem para a viscosidade do algodão, que se tem tornado um problema crescente devido ao seu efeito devastador no processamento e sua produtividade, nomeadamente na fiação *open-end* e de turbina, devidos aos custos de limpeza dos equipamentos impregnados pelos açúcares do algodão (Bradow et al., 2000). Algodões de certas regiões são rejeitados devido aos problemas provocados pelos insectos.

Os açúcares resultam da contaminação das fibras pelos excrementos dos insectos e pelos próprios açúcares constituintes da fibra. Os açúcares são uma mistura complexa de substâncias diferentes, especialmente mono e dioligossacarídeos que contribuem para a rigidez em diferentes graus (Bradow et al., 2000).



### **3.8 – CONSIDERAÇÕES FINAIS**

Com este capítulo deu-se a conhecer um pouco mais a importância histórica da fibra de algodão na produção têxtil, bem como a sua estrutura morfológica e todos os factores influenciadores do seu normal desenvolvimento.

Como o foco principal deste estudo assenta na obtenção de conhecimento válido sobre dados relativos às diferentes propriedades físicas e químicas constituintes da fibra de algodão, foi efectuado neste capítulo uma visão promenorizada sobre cada uma das diferentes características pertencentes a cada uma das propriedades, proporcionando assim um maior e mais promenorizado conhecimento acerca de um dos principais objectivos deste estudo, a qualidade da fibra de algodão.



# Cap.4

## DESCRIÇÃO DA FERRAMENTA

---

Todo o trabalho prático desta dissertação foi efectuado recorrendo à utilização do *Clementine* na sua versão 5.2.

Esta ferramenta foi a escolhida para a realização deste trabalho devido ao facto, entre outros, de estar disponível no departamento de sistemas de informação da Universidade do Minho, e de ser uma ferramenta de referência na área da prospecção de dados, sendo utilizada em vários estudos, alguns deles departamentais, com resultados bastante satisfatórios. Outro dos factores relevantes, reside no facto de que o *Clementine* ser visto como uma ferramenta líder nos trabalhos de prospecção de dados devido ao facto de devolver o máximo possível dos dados com o mínimo de tempo investido (SPSS, 2003).

### 4.1 – INTRODUÇÃO

O *Clementine* é uma ferramenta de prospecção de dados que foi desenhada e desenvolvida com o objectivo de tirar proveito de um recurso subaproveitado, os dados. O *Clementine* permite uma fácil navegação nos dados, utilizando gráficos para localizar relações importantes (Clementine, 1994-1998).

Esta ferramenta, para além de ser de simples utilização e aprendizagem, disponibiliza facilidade de manipulação dos dados, experimentando diferentes combinações de dados e técnicas e testando hipóteses. Essas facilidades ajudam a aplicação dos conhecimentos e perícia dos analistas na descoberta de potenciais valiosas propriedades dos dados (Clementine, 1994-1998).

O *Clementine* também inclui uma modelação avançada ou técnicas de aprendizagem máquina que extraem inter-relacionamentos e regras de decisão a partir dos dados. Isto permite automatizar aplicações tais como previsões, estimativas e classificações e fornecer apoio à decisão (Clementine, 1994-1998).

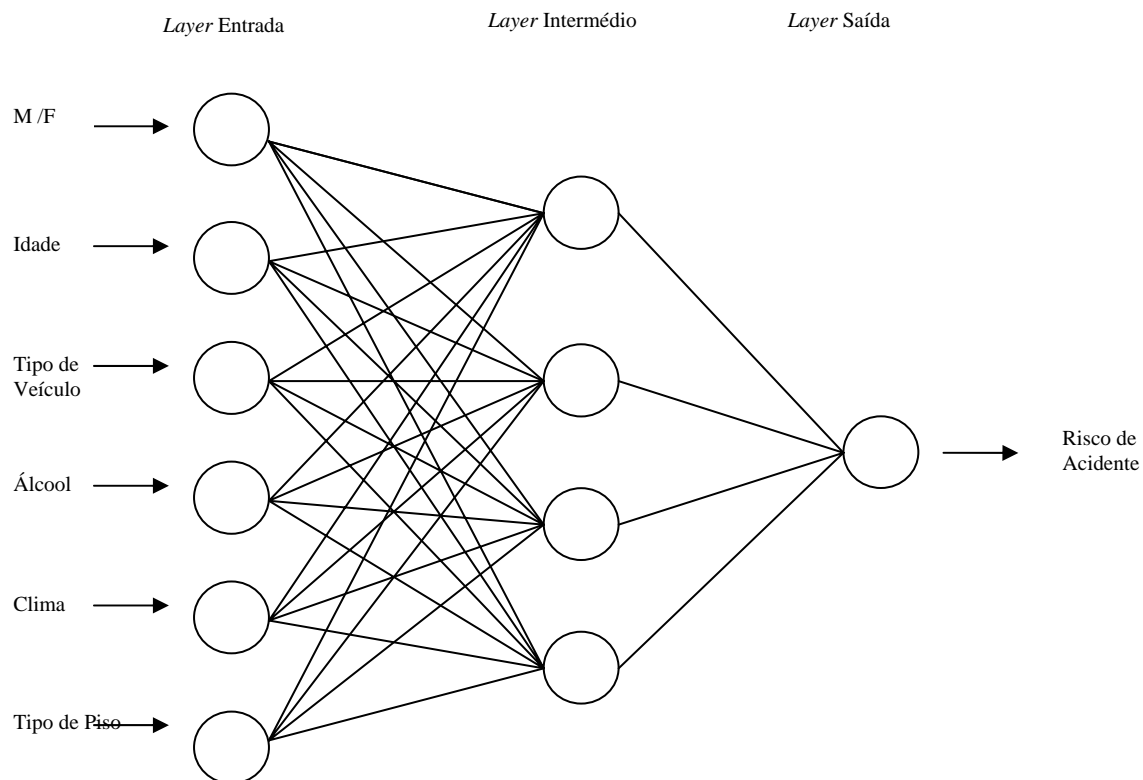


## 4.2 - TÉCNICAS DE MODELAÇÃO

De seguida são apresentadas as técnicas de modelação mais frequentemente utilizadas e disponíveis na ferramenta.

### 4.2.1 - Redes neuronais

As redes neuronais consistem em modelos muito simples, que imitam o funcionamento do sistema nervoso humano (Clementine, 1994-1998). A unidade básica é denominada de “neurónio” e encontram-se normalmente organizados em *layers* ou camadas (figura 3)



**Figura 3** – Rede Neuronal

Os dados de entrada ou *input* são fornecidos à primeira camada e o valor de cada neurónio é propagado para todos os outros neurónios. Os valores dos pesos (*weight*) são alterados durante a transmissão. O resultado é devolvido pelo último neurónio (*output*).



Inicialmente, todos os pesos são aleatórios e as respostas obtidas são provavelmente incorrectas. A rede aprende treinando: resultados conhecidos são constantemente apresentados à rede e comparados com o resultado desta; a informação da comparação é propagada na rede para trás (*backpropagation*), isto se a rede for deste tipo, ajustando gradualmente os pesos. Uma vez treinada, a rede pode ser aplicada a casos em que o resultado é desconhecido.

A ferramenta tem a operação "*Train Net*" que permite criar e treinar uma rede neuronal. Definindo os campos de *input* e de *output* dos dados de treino. A rede neuronal "aprende" a classificar ou prever os valores dos campos de *output* através dos valores dos campos de *input*. Note-se que as regras de decisão para a classificação são internas à rede neuronal, sendo difícil de acompanhar o raciocínio usado no processo de tomada de decisão.

#### 4.2.2 - Indução de regras

Um dos problemas das redes neuronais é a falta de transparência das decisões tomadas pela rede. Visto que o funcionamento interno deste algoritmo se resume a números, torna-se difícil encontrar um raciocínio lógico no processo de tomada de decisão (Clementine, 1994-1998).

As árvores de decisão consistem numa técnica complementar. Trabalhando com a totalidade dos dados ou apenas com um subconjunto, a indução cria uma árvore de decisão que representa regras de como classificar os dados nas várias saídas. A estrutura de árvore e geração de regras são técnicas abertas e explícitas que podem ser navegadas.

Por exemplo, a regra seguinte descreve as características de um dia de ida à praia (figura 4):

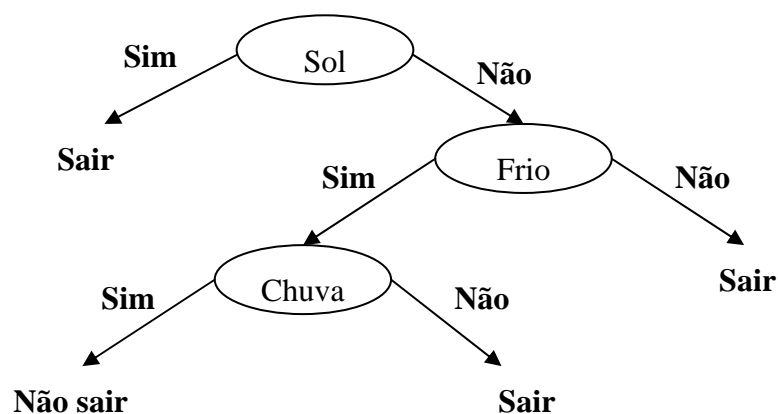


Figura 4 – Árvore de Decisão



Outra vantagem da árvore reside no facto do processo incluir, automaticamente, apenas os atributos que têm realmente importância na tomada de decisão; os outros são descartados. Isto faz com que se obtenha informação útil sobre os dados, sendo possível eliminar os dados que não são relevantes antes de treinar uma rede neuronal. As árvores de decisão podem ser convertidas num conjunto de regras tipo “Se-Então”, dando origem ao conjunto de regras (*rule set*) que, na maioria dos casos, mostra a informação de uma forma mais inteligível (Clementine, 1994-1998).

A estrutura em árvore torna-se muito útil, quando se pretende saber quais os factores que dividem a população nos vários subconjuntos. A estrutura de regras, por seu lado, mostra como um conjunto em particular se relaciona com uma conclusão.

A ferramenta dispõe de duas operações ("Build Rule" e C5.0) que produzem regras sob a forma de árvores de decisão, representando como classificar os dados em diferentes resultados. A estrutura das árvores representa explicitamente o raciocínio da tomada de decisão.

#### 4.2.3 - Redes de Kohonen

As redes de Kohonen são um tipo de rede neuronal que permitem efectuar *clustering*, isto é, dividir a população em segmentos. A unidade básica é o neurónio e estes estão organizados em duas camadas: *layer* de entrada e *layer* de saída que é frequentemente denominado de mapa de saída. Todos os neurónios de entrada estão ligados a todos os neurónios de saída e essas ligações possuem um peso associado (Clementine, 1994-1998). O mapa de saída é uma tabela de neurónios bidimensional sem qualquer ligação entre as células (figura 5).

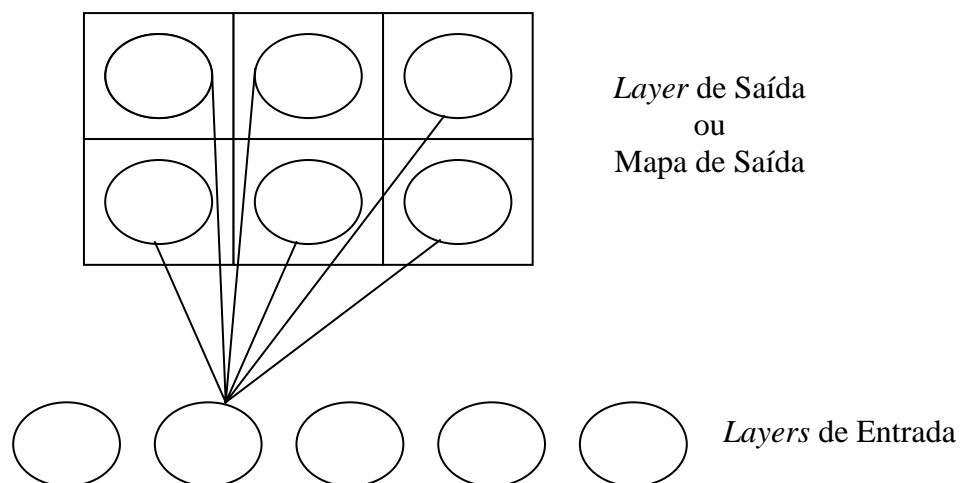


Figura 5 – Redes de Kohonen



Os dados de entrada são fornecidos à camada de entrada e os valores são propagados para o mapa de saída. De seguida, cada neurónio de saída fornece uma resposta. O neurónio de saída que fornecer a melhor resposta é o “vencedor” e constitui a resposta à entrada fornecida.

Inicialmente, todos os pesos são aleatórios. À medida que a rede é treinada, o peso do vencedor é ajustado de forma a responder cada vez melhor, assim como os pesos dos seus vizinhos. Este processo repete-se até as alterações feitas nos pesos serem muito pequenas.

Quando a rede estiver treinada, os registos com características semelhantes aparecerão juntos no mapa de saída, ao contrário dos registos com características distintas, que se encontrarão distantes.

A ferramenta dispõe do algoritmo "*K-Means*" que cria um modelo de agrupamento, isto é, descobre semelhanças nos dados originais e agrupa-os de forma a maximizar a similaridade dentro do grupo e a maximizar a diferença entre grupos.

#### 4.2.4 - Regras de associação

As regras de associação, como o próprio nome indica, associam uma determinada conclusão a um conjunto de condições (Clementine, 1994-1998). Por exemplo a regra:

##### **Basquetebol ← Voleibol & Andebol (173,17.0%,0.84)**

Num contexto desportivo, indica que quem pratica *voleibol* e *andebol*, também pratica *basquetebol*, com um grau de confiança de 84% e representando 17% da população, ou seja 173 registos.

Para descobrir regras, o algoritmo utiliza um método de geração e teste. Inicialmente são criadas regras simples que são validadas mediante o conjunto de dados fornecido. As regras “boas” são armazenadas e todas as regras, após serem submetidas a várias restrições, são “especializadas”. A especialização é o processo que adiciona condições às regras, que são posteriormente validadas. O processo iterativo armazena as regras “melhores” ou “mais interessantes” encontradas. Normalmente o utilizador fornece um limite para o número de conjunções possíveis numa regra e utiliza um conjunto de técnicas baseadas na indexação, que permitem reduzir o tempo de procura. No fim do processo, é criada uma tabela que contém as melhores regras encontradas.



Infelizmente, estas regras, ao contrário das da árvore de decisão, não podem ser utilizadas directamente para fazer previsões, visto existirem diferentes conclusões possíveis. Para obter a árvore de decisão associada a uma determinada regra produzida, é necessário que ela seja tratada separadamente (Clementine, 1994-1998).

A ferramenta dispõe dos algoritmos "GRI" (*Generalised Rule Induction*) e APRIORI que automaticamente encontram associações que existam entre os atributos, explicitando essas regras associativas.

### 4.3 - INTERFACE DO CLEMENTINE

As funcionalidades do *Clementine* estão integradas numa interface de programação gráfica.

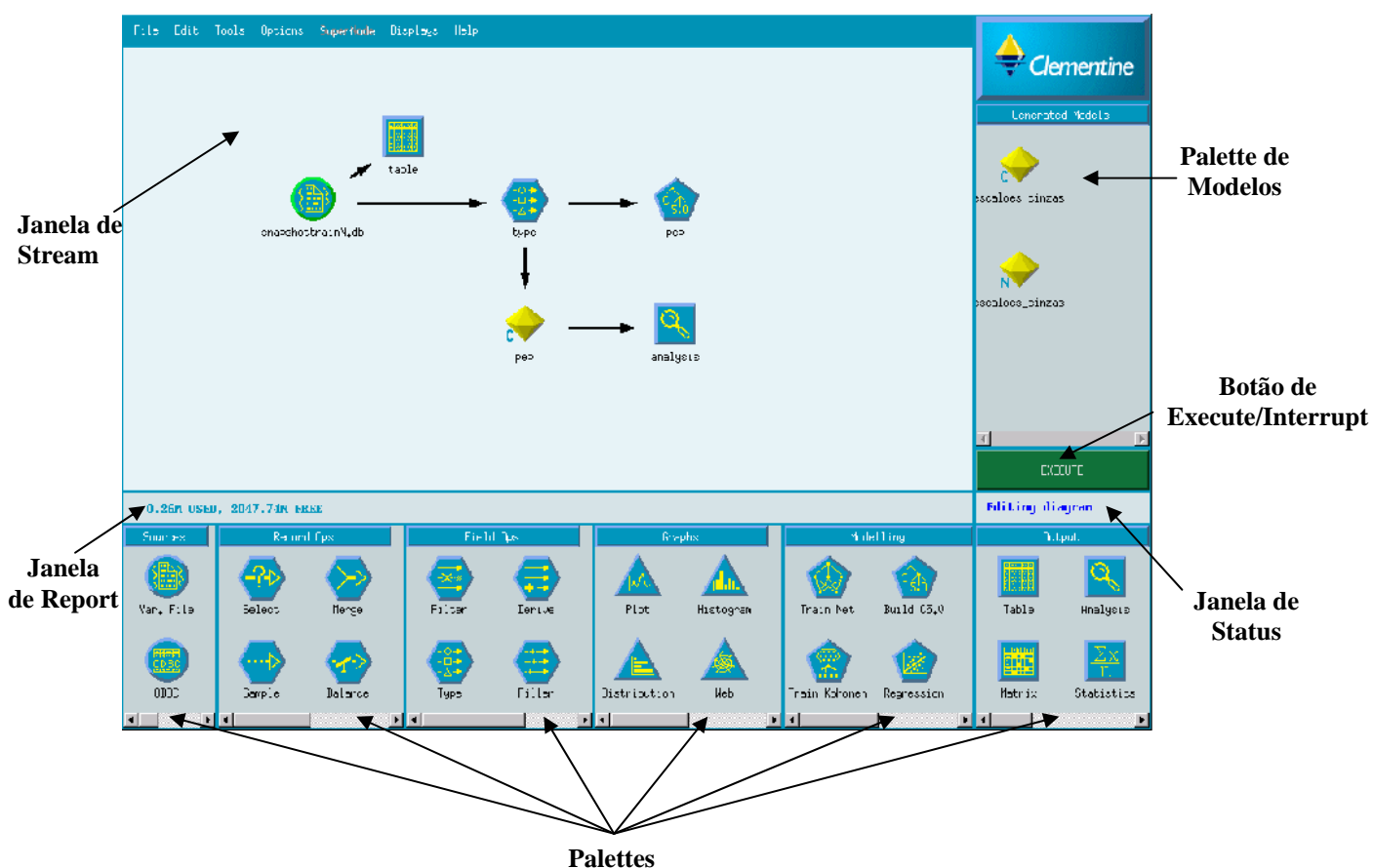


Figura 6 – Janela principal do Clementine com todos os seus componentes.



Na janela principal do *Clementine* (figura 6), encontra-se os seguintes componentes:

- *Stream*: principal área de trabalho do *Clementine*. A programação gráfica envolve a colocação e manipulação de ícones que representam nós de processamento. O conteúdo é designado de diagrama; estes podem ser carregados e gravados de e para uma unidade de armazenamento;
- *Palletes*: contêm famílias de ícones que representam diferentes operações possíveis no *Clementine*. Na programação gráfica, os ícones são seleccionados das *palletes*, colocados na *stream*, conectados uns aos outros e editados;
- *Pallete* de Modelos: esta é uma *pallete* especial que guarda o resultado da modelação;
- Janela de *Report*: a janela de *report* dá-nos um *feedback* do progresso das várias operações. Quando os registos estão a ser lidos, ela mostra-nos o ponto de situação e durante a aprendizagem dos modelos mostra-nos o tempo despendido nesse processo;
- Janela de *Status*: a janela de *status* diz o que o *Clementine* está a fazer ou em que modo está (por exemplo: “*editing diagram*”, “*executing*” ou “*plotting*”). Em algumas circunstâncias dá breves instruções: “*place new filter icon*” ou “*connect to another icon*”;
- Botão de *Execute/Interrupt*: normalmente, este botão está etiquetado com a palavra *Execute*; quando pressionado, executa uma *stream* válida no corrente diagrama. Durante a execução, o botão está etiquetado com a palavra *Interrupt*, permitindo interromper a execução da *stream*, com a opção de abandonar ou continuar o processamento.

#### 4.4 - COMO UTILIZAR O CLEMENTINE

Aproveitando as potencialidades gráficas do *Clementine*, o utilizador desenha um diagrama do que pretende que aconteça aos dados, ligando as várias operações numa *stream*. Cada tipo de operação é representado por ícones, que por sua vez se encontram agrupados em *palettes*. As *palettes* encontram-se no fundo da janela e estão etiquetadas por categorias de acordo com as funções desempenhadas pelos respectivos ícones.

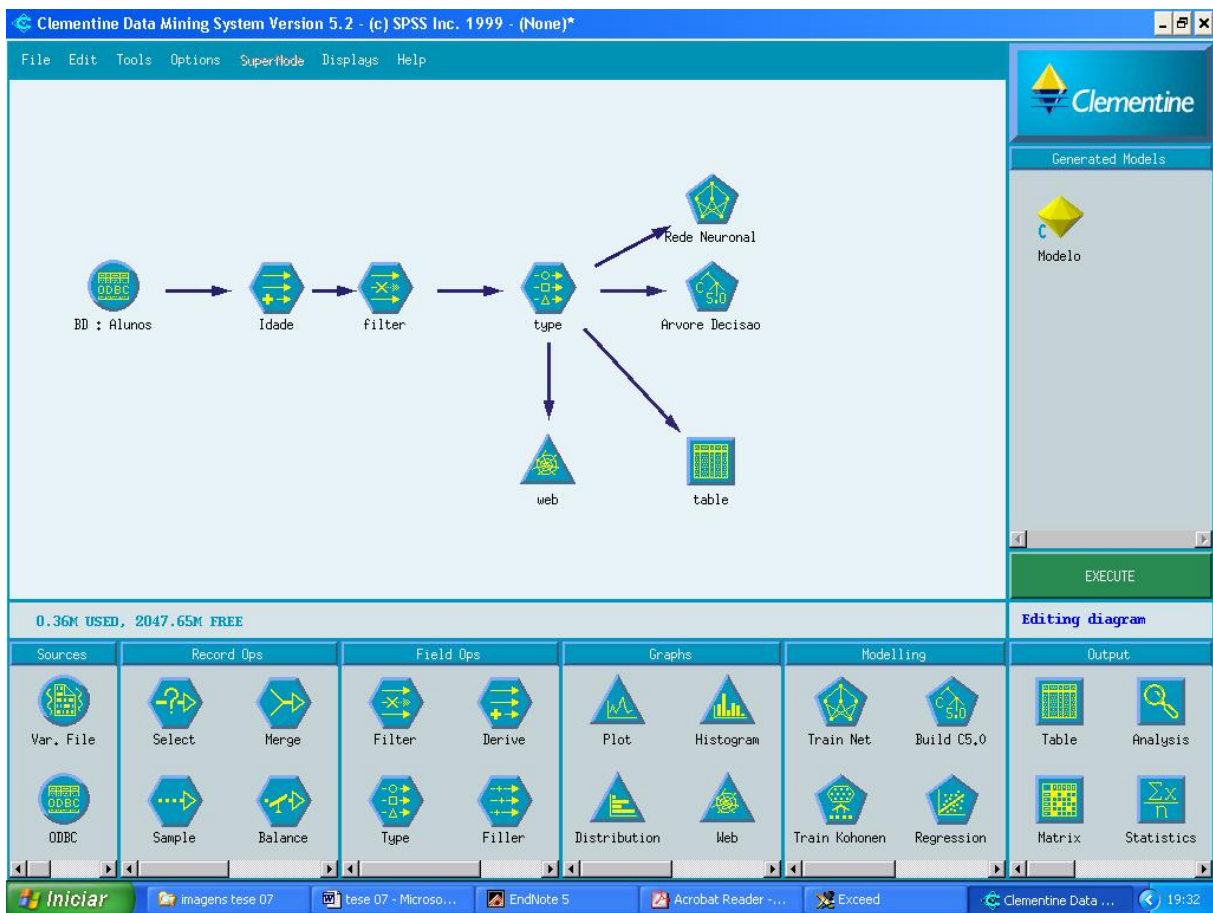


Figura 7 – Exemplo de uma *stream*.

Observando a figura 7, é possível observar a existência de 6 tipos distintos de *palettes*:

- A *palette* dos ícones de origem que permitem fornecer o *input* das *streams*;
- A *palette* dos ícones de operações sobre os registos;
- A *palette* dos ícones de operações sobre os atributos;
- A *palette* dos ícones de visualização gráfica;
- A *palette* dos ícones de modelação;
- A *palette* dos ícones de destino que permitem o armazenamento, visualização e análise do *output* das *streams*.



#### 4.4.1 - Descrição dos ícones

De seguida é dada uma breve descrição dos tipos de ícones existentes nas palletes do *Clementine* e descritas anteriormente:

- Ícones de entrada (*Sources*)

Os *Source Nodes* importam, para o *Clementine*, registos provenientes de ficheiros ASCII ou bases de dados. São sempre o primeiro nó de uma *stream* e não podem receber nenhuma ligação de entrada. Esta família de ícones tem uma forma circular.

- *Fixed File*: lê dados de ficheiros com campos de tamanho fixo;
- *Variable File*: lê dados de ficheiros com campos de tamanho variável, utilizando a informação de caracteres delimitadores para identificar cada campo;
- ODBC: lê dados de uma base de dados relacional utilizando uma vista ODBC;
- *Cache File*: lê dados de ficheiros do tipo *Clementine Cache Format (CCF)*.

- Ícones de Manipulação (*Records Ops e Fields Ops*)

Os *Manipulation Nodes* executam operações nos registos (ex: *select*, *merge* e *sample*) ou nos atributos (ex: *filter*, *derive*, *type* e *history*). Este tipo de ícones suporta tanto as ligações de entrada como as de saída.

Muitos destes ícones recorrem a expressões da linguagem CLEM<sup>6</sup> para criar novos valores. Quando a *stream* é executada, todas as expressões CLEM são compiladas antes de qualquer nó ser executado, havendo uma imediata verificação da sintaxe. Esta família de ícones tem uma forma hexagonal.

- ✓ *Select*: selecciona um subconjunto de registos baseando-se numa condição específica;
- ✓ *Sample*: selecciona um subconjunto contíguo ou aleatório de registos;

---

<sup>6</sup> Linguagem de programação própria do *Clementine*



- ✓ *Merge*: pode receber múltiplas entradas. Combina registos de diferentes entradas para criar uma única saída, utilizando um ou mais atributos em comum;
- ✓ *Balance*: modifica as proporções dos registos de modo a equilibrar as quantidades existentes de cada registo;
- ✓ *Sort*: ordena os registos baseado nos valores das variáveis que lhes são passados como chave;
- ✓ *Aggregate*: agrega os registos que lhe são passados como entrada, para gerar registos de saída sumariados;
- ✓ *Distinct*: remove registos duplicados de acordo com um determinado critério;
- ✓ *Append*: pode receber múltiplas entradas. Faz o *append* de dois ou mais conjuntos de dados;
- ✓ *Filter*: filtra e renomeia campos;
- ✓ *Derive*: adiciona um novo campo ao conjunto de dados. Existem seis tipos de *derive node* que podem ser criados;
- ✓ *Type*: associa tipos de dados a cada atributo do conjunto de dados. Também permite a verificação de tipos;
- ✓ *Filler*: substitui o valor de um determinado atributo por um novo valor;
- ✓ *History*: adiciona novos atributos baseando-se nos valores de atributos anteriores;
- ✓ *Set To Flag*: cria um campo do tipo *flag* baseado no conjunto de valores possíveis para esse campo. As *flags* resultantes podem ser agrupadas.

- Ícones gráficos (*Graphs*)

Os *Graph Nodes* utilizam gráficos para mostrar as propriedades das relações existentes entre os vários atributos. Apenas podem receber uma única ligação de entrada.

Esta família de ícones tem uma forma triangular.

- *Distribution*: permite visualizar a distribuição dos valores de variáveis simbólicas;



- *Histogram*: permite visualizar a distribuição dos valores de variáveis numéricas, através de um histograma;
- *Plot*: permite visualizar a relação entre duas variáveis numéricas, através de um gráfico bidimensional;
- *Web*: permite visualizar a ligação existente entre os valores de varias variáveis simbólicas;
- *Multiplot*: faz o *display* de múltiplas variáveis no eixo do Y's numa única no eixo dos X.

- Ícones de Modelização

Os *Modelling Nodes* utilizam aprendizagem máquina (redes neuronais e regras de indução) e técnicas estatísticas (regressão linear) para prever valores para as variáveis de saída ou grupos de registos semelhantes.

Cada ícone apenas pode receber uma única ligação de entrada.

Esta família de ícones tem uma forma pentagonal.

- *Built Rule*: utiliza árvores de decisão para gerar regras capazes de preverem os valores de uma variável de *output*;
- *Train Net*: cria e treina uma rede neuronal que prevê o valor de uma ou mais variáveis de saída. Por defeito, a rede criada contem *layers* com mais do que um *preceptron* (multi-layer perceptron – MLP), que são treinados através de uma aprendizagem por backpropagation.
- *Train Kohonen*: cria e treina uma rede de *Kohonen* a fim de dividir os registos em *clusters*;
- *Built C5.0*: gera uma árvore de decisão ou um conjunto de regras capazes de prever o valor da variável de saída;
- *Regression*: recorre à regressão linear para produzir a “melhor” equação, capaz de modelar uma variável de *output* numérica, tendo em conta várias variáveis de entrada numéricas;



- GRI: cria um conjunto de regras independentes que descrevem o conjunto de dados. O GRI generaliza as regras de associação e a sua funcionalidade é por vezes referenciada como sendo o “descobridor” de regras de associação. Suporta, tanto as variáveis numéricas como as simbólicas, como dados de entrada e apenas variáveis simbólicas como saída. Este modelo produz regras que não estão refinadas;
  - *Apriori*: cria um conjunto de regras de associação independentes que descrevem o conjunto de dados. O algoritmo de modelização do *apriori* faz a descoberta de regras de indução. O *apriori* apenas suporta variáveis simbólicas e produz regras que não estão refinadas;
  - *Train Kmeans*: produz um modelo que faz *clustering* utilizando um determinado número de *clusters*.
- Ícones de Saída (*Output*)

Os *Output Nodes* lêem os dados e guardam uma tabela ou um *report* num ficheiro ou numa base de dados. Apenas podem ter uma única ligação de entrada.

Esta família de ícones tem uma forma quadrada.

- *Table*: faz o *display* ou grava os dados de uma forma tabular;
- *Matrix*: faz o cruzamento de duas variáveis simbólicas; opcionalmente, mostra valores associados para uma determinada variável numérica. Pode também relacionar um conjunto completo de *flags* ou de variáveis numéricas;
- *Output File*: guarda os dados num ficheiro com tamanhos de campos variáveis ou em ficheiros *Clementine Cache Format (CCF)*;
- *Report*: produz um *report* de acordo com o formato indicado;
- *Statistic*: elabora uma análise estatística das variáveis numéricas;
- *Analisis*: analisa a performance dos modelos gerados;
- *Set Globals*: calcula os valores a serem desenvolvidos pelas funções do *Clementine*;



- *Set Quality*: produz um *report* no qual estão determinadas todas as variáveis recebidas que contêm valores inválidos (nulos ou em branco);
- *ODBC Output*: guarda os dados numa tabela da base de dados relacional usando o ODBC;
- *Excel*: invoca o *Microsoft Excel*.

#### 4.4.2 - Criação da *Stream*

O utilizador selecciona ícones das *palettes* e coloca-os na área de desenho, estes passam então a chamarem-se nós e são ligados através de setas que indicam o sentido do fluxo da informação na *stream*.

Normalmente a *stream* começa com um nó de origem, seguindo-se os nós de manipulação dos dados (usados para transformar os dados para o formato pretendido, antes de estes serem listados, armazenados ou modelados) e finalmente nós terminais. Os destinos dos dados podem ser *reports*, gráficos, técnicas de modelação ou tabelas.

Como exemplo, a *stream* representada na figura 7 contém um nó de origem (ícone circular), vários nós de manipulação (ícones hexagonais) e vários nós terminais (ícones triangulares para visualização gráfica, ícones quadrados para outras formas de visualização e armazenamento e ícones pentagonais para a modelação). O processo de modelação produz modelos que são representados por um ícone com a forma de um diamante na *palette* de modelos. Estes ícones também podem ser utilizados na *stream*.

Cada nó pode ser editado, utilizando uma caixa de diálogo, para serem especificados detalhes sobre as operações a serem executadas. Quando a *stream* é executada, a informação flui no sentido das setas desde a origem e até ao destino.

#### 4.4.3 - Super-nós

Uma das razões pela qual o *Clementine* é tão simples de aprender, consiste na clara definição das funções de cada nó. Este facto, no entanto, pode levar a que uma *stream* se torne muito complexa se uma grande sequência de nós for necessária (Clementine, 1994-1998).



Para solucionar este problema, pode-se optar por dividir a *stream* em várias. A primeira cria um ficheiro que será o *input* da segunda e assim sucessivamente. Esta solução é eficaz, mas implica uma constante limpeza e carregamento de *streams*.

Em alternativa, podem ser utilizados super-nós. Os super-nós não se encontram em nenhuma das *palettes* e são representados por uma estrela. A sua função é permitir o agrupamento de fragmentos de *stream* (constituídos por vários nós) num só nó. A este agrupamento dá-se o nome de encapsulamento.

A utilização de super-nós torna-se vantajosa visto que mantém a *stream* clara e manuseável, permite uma fácil compreensão (mesmo para outros utilizadores) tendo em conta que os super-nós têm um baixo nível de detalhe e podem ser exportados para bibliotecas e reutilizados noutras *streams*.

Podemos ter diferentes tipos de super-nós:

- Super-nós de entrada: contêm um nó de entrada;
- Super-nós terminais: contêm pelo menos um nó de entrada;
- Super-nós de manipulação: apenas contêm nós de manipulação.

Existem, no entanto, algumas limitações na utilização deste tipo de nó e no processo de encapsulamento:

- Tem que existir um caminho entre os dois nós escolhidos para o encapsulamento;
- Um super-nó não pode conter um nó de entrada e um nó de saída em simultâneo;
- Um super-nó não pode incluir nós do tipo *merge* ou *append*;
- Um super-nó não pode conter outro super-nó;
- Nenhum dos nós a ser encapsulado pode ter múltiplas conexões de saída, como por exemplo, uma bifurcação na *stream*.

Esta ultima restrição não se aplica aos nós terminais. Neste caso, todos os nós terminais são incluídos no super-nó.



#### **4.5 - CONSIDERAÇÕES FINAIS**

A ferramenta *Clementine*, integra um conjunto de algoritmos de modelação e aprendizagem, com uma interface de programação visual que torna compreensível todo o processo de extracção de conhecimento a partir de diversas fontes de dados.

O utilizador elabora diagramas ou *streams* onde define as sequências de operações que deseja realizar sobre os dados (o processo de extracção de conhecimento). Estas operações são representadas por ícones que estão agrupados de acordo com o seu tipo de funcionalidade.



# Cap.5

## TRABALHO PRÁTICO

---

Neste estudo, tal como em todos na generalidade, o primeiro passo consistiu na obtenção de uma ideia clara acerca do problema que se pretende resolver, após o qual foi efectuada uma análise de toda a informação recolhida e armazenada na base de dados em estudo. Esta análise foi conseguida através do uso de técnicas de visualização existentes no Clementine.

O objectivo desta dissertação consistiu na utilização de técnicas de prospecção de dados na obtenção de regras e modelos com um grau de certeza elevado através da utilização do *Clementine*. Possibilitando a descoberta e obtenção de relações existentes entre propriedades físicas e químicas das fibras de algodão, de que forma se relacionam e qual a influencia entre cada uma delas.

A potencialidade de visualização dos resultados através desta ferramenta surge como uma técnica bastante útil, permitindo a análise gráfica das possíveis relações. A representação gráfica proporciona a visualização de perspectivas que normalmente não são possíveis visualizar. As técnicas de visualização proporcionam à percepção humana vantagens no método de análise (Dias et al., 2002).

### 5.1 - SELECÇÃO E PREPARAÇÃO DAS BASES DE DADOS

Numa primeira fase, iniciou-se a construção do Armazém de Dados propriamente dito. O processo de ECBD pode ser assim dividido em dois níveis: um primeiro nível, do qual fazem parte os passos relativos à selecção e pré-processamento dos dados, em que a partir do ambiente operacional são capturados e transformados os dados que depois constituem um Armazém de Dados; e um segundo nível, em que a partir dos dados no Armazém de Dados podem ocorrer novas transformações e em que é efectuada a captura de padrões de relacionamento (Baptista et al., 2000).

O objectivo do Armazém de Dados é o de proporcionar as infra-estruturas necessárias de armazenamento, acesso e manipulação de dados, organizando os dados operacionais de forma a otimizar o processo de extracção de conhecimento. Assim, o Armazém de Dados funciona como a plataforma de suporte para as aplicações de prospecção de dados (Baptista et al., 2000).



No entanto, independentemente do tamanho da tabela referente aos dados químicos ser menor, efectuou-se a divisão das base de dados, para treino e para validação, numa proporção de 60% por 40%. Para se proceder ao início do trabalho de prospecção de dados tornou-se necessário agrupar os dados químicos existentes, com os todos os dados físicos recolhidos. É de notar o menor número de dados químicos existentes, o que poderá levar a que os resultados finais não atinjam os níveis de certeza desejados.

Todo o armazém de dados é mantido em uma base de dados *Access*. A tabela relativa aos dados físicos é composta por 5936 registos de várias origens: USA, Costa do Marfim, Moçambique, Sudão, Rússia, Brasil, Egipto, Zimbabwe, Camarões, Madagáscar, Uganda, China, Benin e Tchad. No entanto e devido ao avultado custo financeiro que acarreta a elaboração de testes químicos, bem como da morosidade dos referidos testes, a todos os fardos de algodão provenientes das diferentes origens, só foi possível efectuar testes sobre oito dessas origens sendo criada uma nova tabela referente aos dados químicos, com um total de 197 registos, sendo estes referentes a fardos provenientes do Brasil, Rússia, Egipto, Zimbabwe, Camarões, Madagáscar, Marrocos e Uganda. A figura 8 mostra um excerto da tabela referente aos dados das propriedades físicas das fibras de algodão.

ID_fisicos	origem	fardo	micrinaire	UHM	SL50	SL2,5	UI	UR
1	usa	0	4.2	34.3	25.7	34.3	91.2	74.9
2	usa	1	3.8	36.3	21.6	36.1	87.6	60
3	usa	2	4.2	33.5	24.6	33.3	91.6	73.8
4	usa	3	3.9	33.8	21.1	33.8	86.5	62.6
5	usa	4	4.2	34	27.7	33.8	93.5	81.3
6	usa	5	4.1	35.3	26.9	35.1	93.3	76.9
7	usa	6	4.4	36.1	27.2	35.8	90.6	75.3
8	usa	7	4.5	33.3	22.4	33	86.3	67.5
9	usa	8	4.1	34	23.6	34	90.1	70
10	usa	9	4	33.8	20.3	33.3	87.5	61.1
11	usa	10	4	35.8	25.7	35.8	89.8	70.3
12	usa	11	4.1	34.3	22.9	34.8	89.3	65.5
13	usa	12	3.8	34	23.1	34	87.8	67.6
14	usa	13	4.1	35.3	27.9	35.1	93.5	79.6
15	usa	14	4.1	35.1	23.6	34.5	89.5	68.5
16	usa	15	4.5	34.5	25.4	34.3	93.1	73.8
17	usa	16	4.6	35.3	26.2	34.8	90.4	75.1
18	usa	17	4.5	35.6	26.4	35.3	91.6	74.5
19	usa	18	4.1	33.5	22.4	33.5	88.9	66.2
20	usa	19	3.8	34.5	26.4	34.3	91.3	77.2
21	usa	20	4.3	33.8	23.4	33	90	70.7
22	usa	21	4.4	36.6	25.9	36.8	89.3	70.9

Figura 8 – Excerto da tabela de dados referentes às propriedades físicas.



O processo de junção das duas tabelas iniciais, contendo os registos referentes aos dados físicos e químicos, bem como da criação das duas novas tabelas com os registos agrupados, está descrito na figura 9. Como resultado foram obtidas duas novas tabelas, sendo uma designada de *Treino* e a outra de *Validacao*. A tabela para treino vai ser usada para criar e refinar escalões que serão gerados, bem como, para aplicar os modelos considerados como os mais adequados. Após o treino, será efectuada a validação dos modelos obtidos, utilizando a tabela de validação.

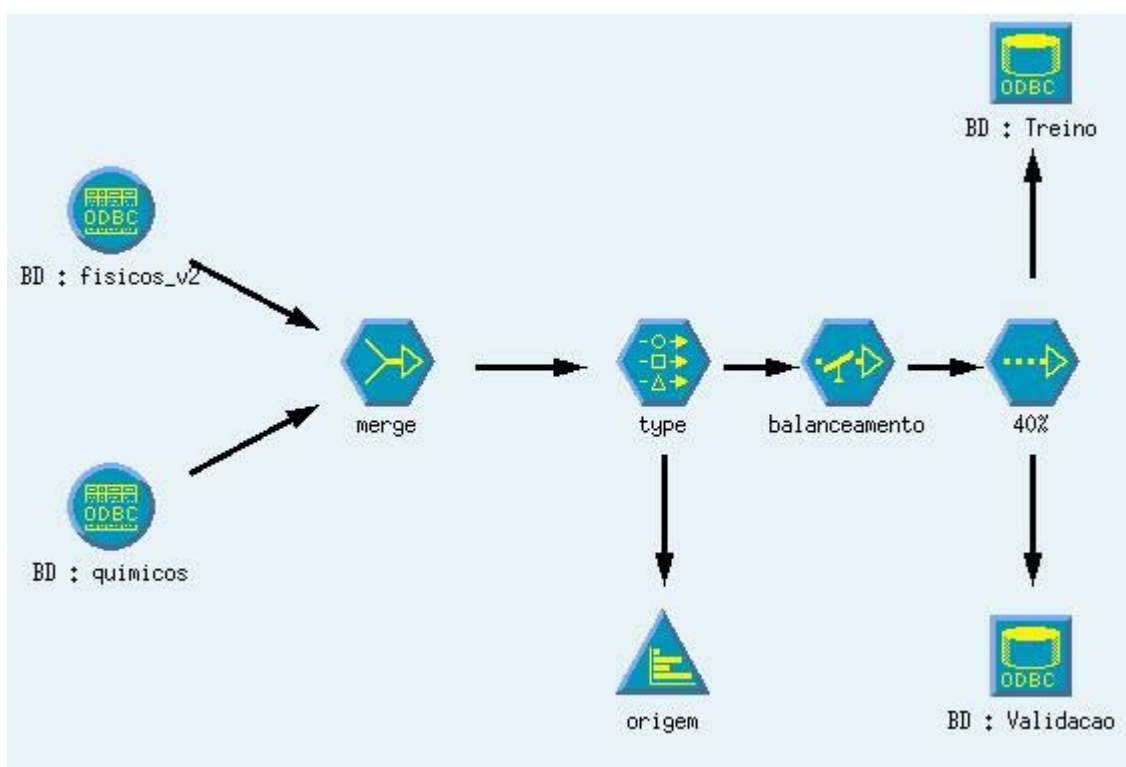


Figura 9 – Stream separacao.str

A característica fundamental desta *stream* foi a necessidade de aplicação de um gráfico de distribuição dos registos por origem, após a junção de todos os registos efectuada através de uma operação de *merge* e assumindo como chave primária o campo origem. Com um gráfico de distribuição gerou-se um ícone de balanceamento (*balance*) para assim garantir a presença de registos relativos a todas as origens nas duas tabelas geradas, *Treino* e *Validação*.



## 5.2 - CRIAÇÃO DE ESCALÕES DAS PROPRIEDADES QUÍMICAS EM ESTUDO

As propriedades químicas em estudo consistem na percentagem de cinzas (%cinzas), Potássio (K), o Magnésio (Mg), o Fósforo (P), Cálcio (Ca) e a percentagem de materiais extraídos (ME). Assim nesta parte do estudo pretende-se prever o comportamento das várias variáveis de entrada para as variáveis de saída definidas.

É importante salientar que as variáveis Potássio, Magnésio, Fósforo e Cálcio, são obtidas após análise das cinzas dos diferentes fardos de algodão, provenientes das origens em estudo. Os elementos constituintes das cinzas constituem um espectro de propriedades muito mais alargados, no entanto estes são os elementos existentes em maior percentagem e comuns a todos os algodões, assim sendo, dentro desse conjunto alargado de elementos, são considerados como os mais relevantes.

Analisando os registos existentes na base de dados, referentes às propriedades químicas, verificou-se que são definidas por um conjunto alargado de valores, o que tornou necessário a criação de escalões que tornassem possível aplicação de modelos necessários para o desenvolvimento do estudo. Assim e através da *stream cria\_esc\_quim.str* (figura 10), utilizando histogramas (*histogram*), foram criados vários nós de derivação (*derive nodes*), cada um deles referente à propriedade analisada em cada histograma. Como resultado final foi obtidas novas tabelas com os novos campos: *esc\_%cinzas*, *esc\_ME*<sup>7</sup>, *esc\_K*, *esc\_Mg*, *esc\_P* e *esc\_Ca*.

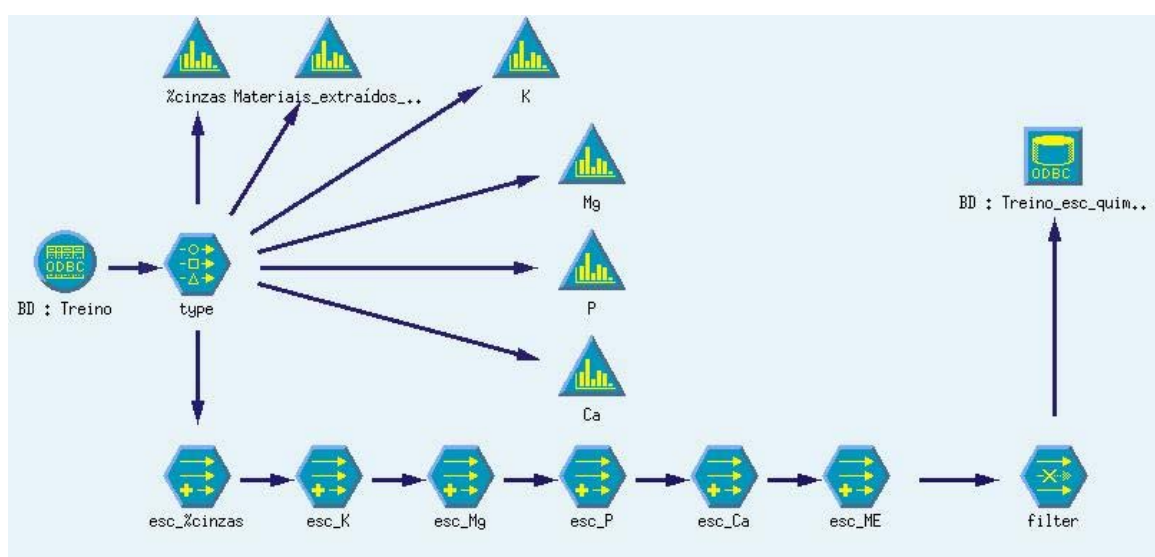


Figura 10 – *Stream cria\_esc\_quim.str*

<sup>7</sup> Materiais Extraídos



A *stream* aplicou-se, quer aos dados de treino, quer aos dados de validação, tendo como resultado final, duas novas tabelas de nome: *treino\_esc\_quimicos* e *validacao\_esc\_quimicos*.

Os escalões criados relativamente às propriedades químicas podem ser analisados nas tabelas 11,12,13,14, 15 e 16.

Escalões	% Cinzas
A	< 1.23
B	[ 1.23;1.37 [
C	[ 1.37;1.46 [
D	>= 1.46

**Tabela 11** – Escalões de % cinzas gerados.

Escalões	% ME
A	< 7.28
B	[ 7.28;7.82 [
C	[ 7.82;10.02 [
D	>= 10.02

**Tabela 12** – Escalões de % ME gerados

Escalões	K
A	< 55.64
B	[ 55.64;67.04 [
C	[ 67.04;69.57 [
D	>= 69.57

**Tabela 13** – Escalões de K gerados

Escalões	Mg
A	< 4.58
B	[ 4.58;5,00 [
C	[ 5,00;6.32 [
D	>= 6.32

**Tabela 14** – Escalões de Mg gerados.

Escalões	P
A	< 2.53
B	[ 2.53;3,00 [
C	[ 3,00;4.30 [
D	>= 4.30

**Tabela 15** – Escalões de P gerados.

Escalões	Ca
A	< 15.02
B	[ 15.02;17.49 [
C	[ 17.49;23.29 [
D	>= 23.29

**Tabela 16** – Escalões de Ca gerados.

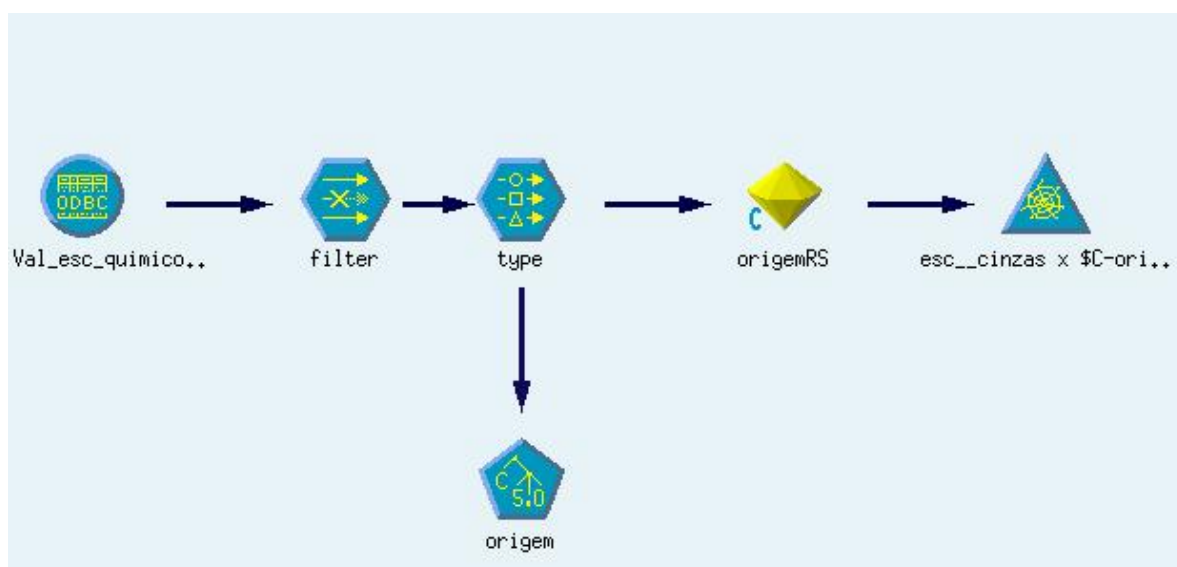
A criação dos escalões, tal como constam nas tabelas acima representadas, foram sustentados na aplicação de gráficos de distribuição (*Distribution*) sobre cada nova variável, verificando desta



forma a distribuição uniforme de ocorrências dentro de cada escalão, para assim não influenciar o comportamento das novas variáveis.

### 5.3 - RELAÇÃO ENTRE AS PROPRIEDADES QUÍMICAS E A ORIGEM DA FIBRA

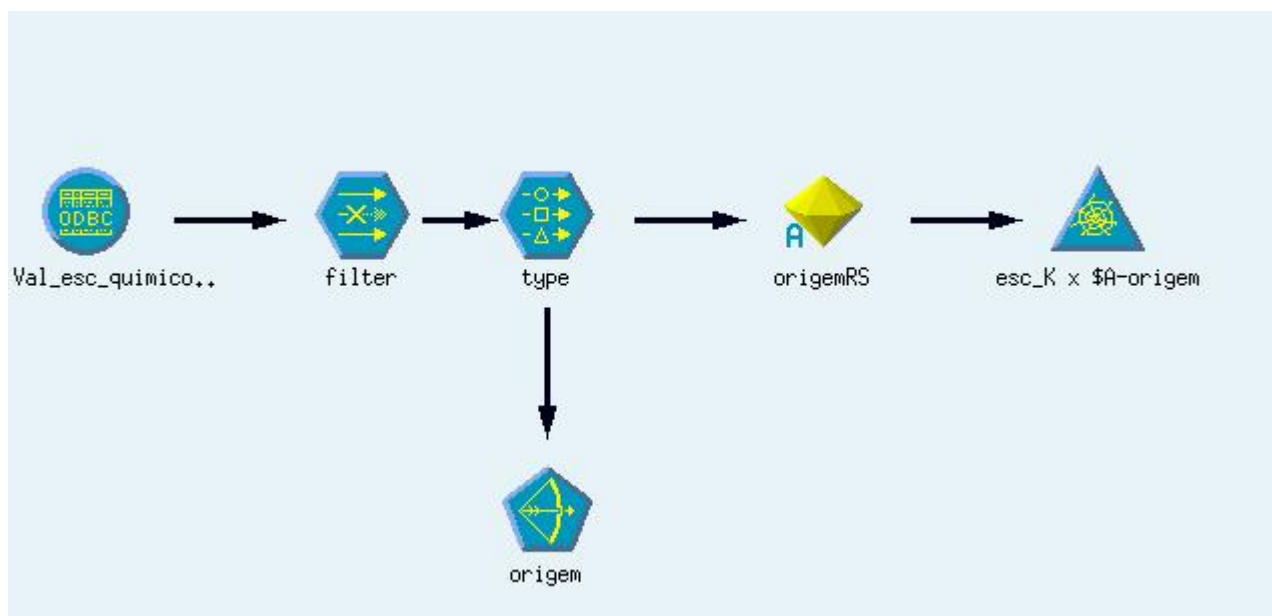
De seguida, e para melhor compreensão da forma como as fibras provenientes das diferentes origens estão relacionadas com as propriedades químicas em estudo, gerou-se a *stream relacao.str* (figura 11).



**Figura 11** – *Stream relacao.str* (utilizando o modelo C5.0)

Nesta *stream* foi utilizado o modelo C5.0 (indução de regras), gerando um conjunto de regras sob a forma de árvores de decisão, capazes de prever o valor da variável de saída “origem”, tendo unicamente como variáveis de entrada as relativas aos dados químicos, sendo todos os restantes filtrados através de um ícone de manipulação (*filter*). Posteriormente, através de gráficos do tipo *Web*, que permitem efectuar uma visualização gráfica dos resultados obtidos, visualizaram-se de forma explícita as relações entre o campo origem e as variáveis químicas, separadamente.

De forma a confrontar e refinar os possíveis resultados obtidos através da utilização de diferentes modelos existentes na ferramenta, efectuou-se a aplicação na mesma *stream* do modelo GRI (*Generalised Rule Induction*) para induzir um conjunto de regras de associação explícitas e quantificadas (figura 12).



**Figura 12** – *Stream relacao.str* (utilizando o modelo GRI)

O resultado deste modelo é dado sob forma de regras de associação, indicando para cada uma delas a sua cobertura e confiança (quantas ocorrências são cobertas pelas condições das regras e a proporção destas em que a consequência é também verdadeira). Na configuração deste modelo foi possível definir o número máximo de regras a mostrar, bem como a possibilidade de gerar apenas regras que tenham uma confiança superior a 50%. O resultado da aplicação do modelo GRI aparece sob a forma de regras que podem ser observadas em anexo I.

Tomando como exemplo a primeira regra gerada:

**origem == brasil <= esc\_Ca == B & esc\_ME == B (175:13.566%, 1.0)**

Esta regra significa que se o escalão de cálcio e de materiais extraídos da fibra correspondem ao escalão B, então também correspondem à origem Brasil, esta regra tem uma cobertura de 13.57% e uma confiança de 100%.

Através da visualização dos gráficos *Web*, verificou-se que o resultado final é exactamente igual com a aplicação dos dois modelos. O resultado dos gráficos pode ser visto através da análise do anexo I.



De seguida (tabela 17) para melhor compreensão, apresenta-se uma síntese dos resultados obtidos, onde é de salientar que apenas foram consideradas como relevantes as ligações fortes (linhas mais grossas) visualizadas nos gráficos colocados em anexo I.

Origem	Escalões					
	% Cinzas	ME	K	P	Ca	Mg
Brasil	C, D	B	C	B	B	A
Zimbabwe	A, B	B	D	B	A	B
Rússia	B, C	A	B	C	C	C
Camarões	B, C	C	D	A	B	A
Madagáscar	C, D	B	C	B	C	C
Uganda	B, C, D	A	B	C	B	A
Marrocos	A	D	A	D	D	D
Egipto	A, B	A	A	D	D	D

Tabela 17 – Escalões químicos por origem.

#### 5.4 - IMPORTÂNCIA DAS PROPRIEDADES FÍSICAS

Procedeu-se à filtragem de campos, relativos às propriedades físicas, cuja importância seja baixa ou nula na obtenção de resultados. Assim, através da elaboração de uma *stream* de nome *importancia.str*, tomando como variável de *output* cada uma das variáveis correspondentes a uma das propriedades químicas em estudo, aplicou-se um modelo baseado em criação de regras tendo como base redes neuronais, na qual se gerou um nó de selecção, onde automaticamente se efectuou a filtragem de campos cuja relevância para o estudo a efectuar se revelou como nula ou inferior a 50%. Na figura 13 é exemplificado um destes processos, tendo como variável de saída os escalões de material extraído (esc\_ME).

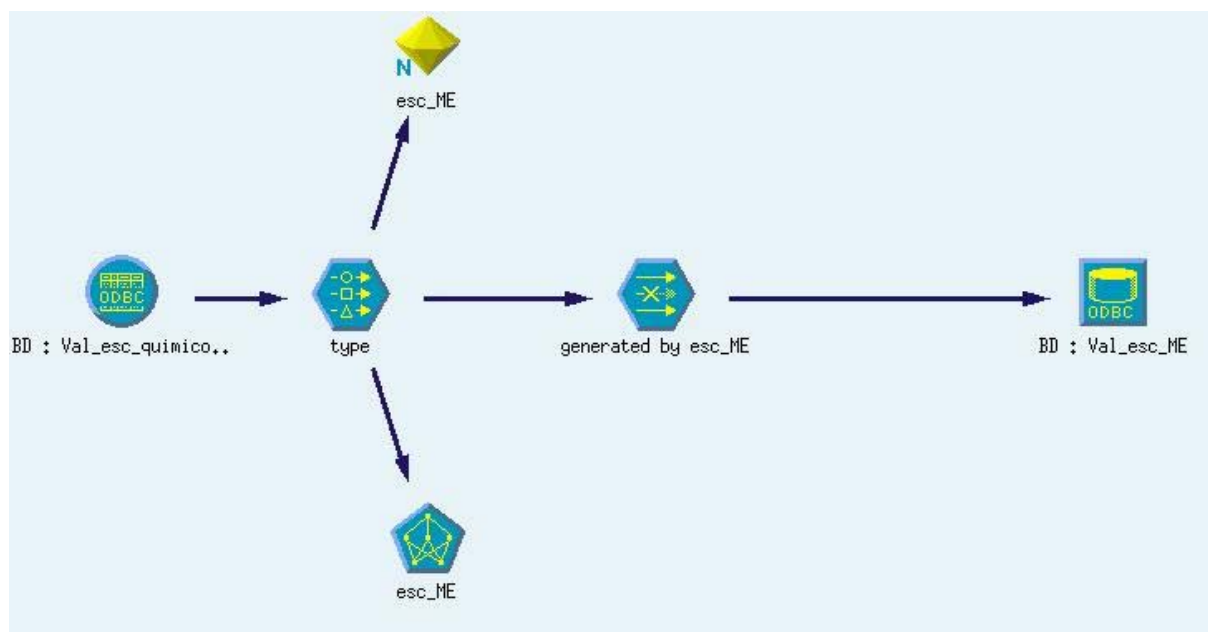


Figura 13 – Stream importancia.str

Com a aplicação da *stream* da figura 13 resultaram duas novas tabelas para cada variável de saída (ex: *Treino\_esc\_ME* e *Val\_esc\_ME*).

## 5.5 - RELAÇÕES ENTRE AS PROPRIEDADES QUÍMICAS E FÍSICAS

Após a verificação da importância das variáveis de *input* (propriedades físicas) para cada variável de *output* (propriedades químicas) a analisar, iniciou-se uma nova fase no processo de prospecção dos dados, a referente à descoberta das possíveis relações entre as propriedades físicas relevantes e cada uma das propriedades químicas, de forma individualizada.

### 5.5.1 - %Cinzas (Percentagem de cinzas)

Tendo como variável de *output* a propriedade química “*esc\_%cinzas*” procedeu-se à criação de escalões sobre as propriedades físicas relevantes, encontradas através da aplicação da *stream* da figura 13. Foram criados escalões sobre três propriedades físicas relevantes, o grau de amarelo (+b), o micronaire e a reflectância (RD). Todo este processo está descrito na figura 14.

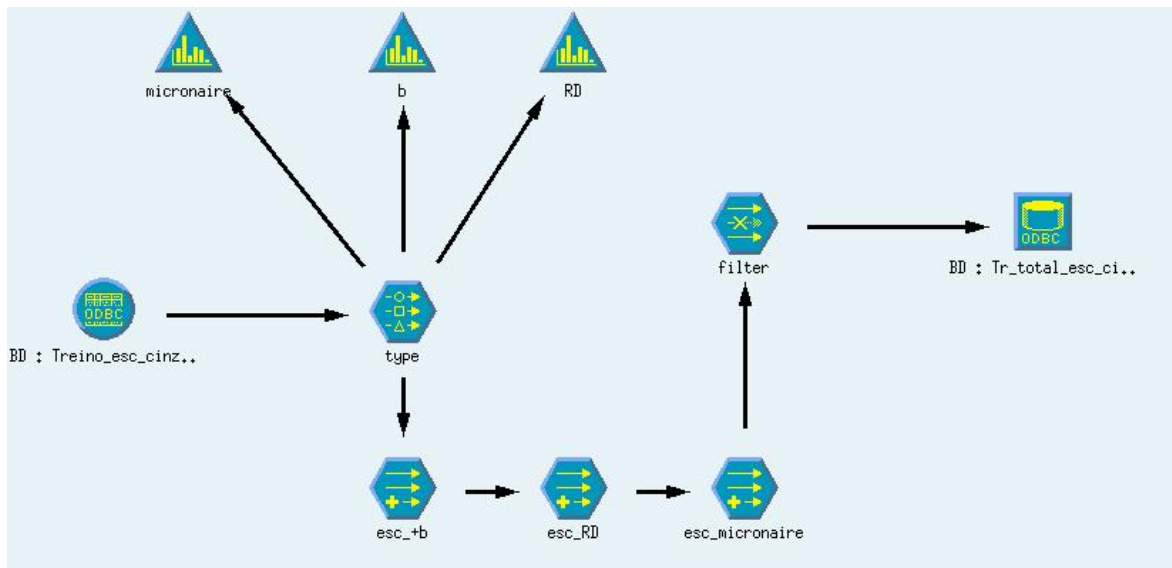


Figura 14 – Stream esc\_total.str

A stream da figura 14 foi aplicada quer para teste, quer para validação, resultando duas novas tabelas de nome “tr\_total\_esc\_cinzas” e “val\_total\_esc\_cinzas”.

As tabelas resultantes da stream da figura 14 consistem então no ponto de partida para o estudo efectuado relativamente à propriedade química de output.

Tal como nos escalões criados anteriormente para as propriedades químicas, a criação das novas variáveis, esc\_+b, esc\_micronaire e esc\_RD, foi sustentada na aplicação de histogramas e gráficos de distribuição sobre cada nova variável criada, verificando desta forma a distribuição uniforme de ocorrências dentro de cada escalão,

Os escalões criados relativamente às propriedades físicas em estudo podem ser analisados nas tabelas 18,19 e 20.

Escalões	+b
A	< 8.96
B	[ 8.96;9.57 [
C	[ 9.57;10.46 [
D	>= 10.46

Tabela 18 – Escalões de +b grados.

Escalões	Micronaire
Muito Fina	< 3.50
Fina	[ 3.50 ; 3.90 [
Média	[ 3.90 ; 4.90 [
Grossa	>= 4.90

Tabela 19 – Escalões de micronaire grados.

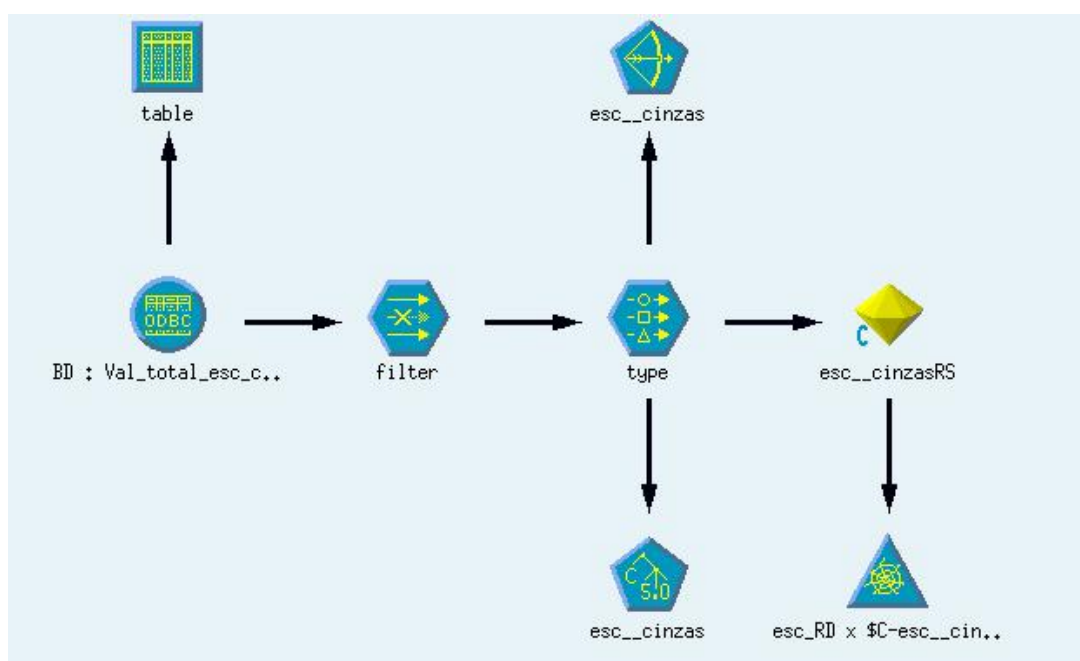


Escalões	RD
A	< 71.33
B	[ 71.33 ; 75.60 [
C	[ 75.60 ; 78.75 [
D	>= 78.75

Tabela 20 – Escalões de RD gerados.

### 5.5.1.1 – RESULTADOS OBTIDOS

Após a criação dos escalões relativos às propriedades físicas, procedeu-se à determinação das relações entre as propriedades físicas relevantes e a propriedade química “% cinzas”. Para tal foi elaborada uma nova *stream* de nome *analise\_relacoes.str*. Tal é demonstrado na figura seguinte.

Figura 15 – Stream *analise\_relacoes.str*

Nesta *stream* foram aplicados ambos os modelos, C5.0 e GRI, após os quais foram geradas um conjunto de regras, que serviram de suporte à visualização dos resultados efectuada com a aplicação de gráficos *Web*.



Os resultados estão relatados em anexo II.

Tomando como exemplo a primeira regra gerada através da aplicação do modelo GRI:

**esc\_cinzas == A <= esc\_b == D & esc\_RD == A (202:15.659%, 0.68)**

Esta regra traduz que, quando uma fibra tem o índice de micronaire situado no escalão D e um índice de reflectância no escalão A, então a percentagem de cinzas dessa mesma fibra encontra-se no escalão A. Esta regra tem uma cobertura de 15.66% e uma confiança de 68%.

De seguida é demonstrada uma síntese dos resultados através da elaboração de uma tabela, cruzando os dados contidos na tabela 17 com a análise aos gráficos em anexo II.

Escalões de % Cinzas	Escalões			Origem
	+b	Micronaire	RD	
A	D	Fina, Média	A	Zimbabwe, Egipto, Marrocos
B	C	Média	B, D	Rússia, Camarões, Uganda, Zimbabwe, Egipto, Marrocos
C	-	Fina	D	Brasil, Madagáscar, Uganda, Rússia, Camarões
D	A, B	Média	B, C, D	Brasil, Madagáscar, Uganda

**Tabela 21** – Matriz escalões de %cinzas vs. propriedades físicas relevantes.

### 5.5.2 - Ca (Cálcio)

Tomando como variável de *output* a propriedade química *Ca* (cálcio) e com a aplicação da *stream* da figura 13 obtiveram-se quatro propriedades físicas relevantes: UI (índice de uniformidade), ST (resistência), RD (reflectância) e +b (grau de amarelo). Como os escalões relativos às propriedades RD e +b foram criados anteriormente, foi somente necessário criar escalões relativos às restantes duas propriedades. Com a *stream esc\_total.str* (figura 14) geraram-se dois novos campos de nome, *esc\_UI* e *esc\_ST*, mais a junção dos campos *esc\_RD* e *esc\_+b*, obtendo-se duas novas tabelas de nome *tr\_total\_esc\_Ca* e *val\_total\_esc\_Ca*.



Os escalões criados relativamente às propriedades físicas em estudo podem ser analisados nas tabelas 18, 20, 22 e 23.

Escalões	UI
Médio	< 82
Elevado	[ 82 ; 85 [
Muito Elevado	>= 85

**Tabela 22** – Escalões de UI gerados.

Escalões	ST
A	< 29.22
B	[ 29.22 ; 32.12 [
C	[ 32.12 ; 37.60 [
D	>= 37.60

**Tabela 23** – Escalões de ST gerados.

### 5.5.2.1 - RESULTADOS OBTIDOS

Com a aplicação da *stream analise\_relacoes.str* (figura 15), obtiveram-se regras através da aplicação dos modelos C5.0 e GRI, após as quais se efectuaram a visualização dos resultados através de gráficos *Web*. Estes resultados estão demonstrados no anexo III.

Para verificarmos o grau de confiança das regras obtidas, vamos tomar como exemplo a segunda regra gerada através da aplicação do modelo GRI:

**esc\_Ca == B <= esc\_ST == A (323:25.039%, 0.85)**

Esta regra traduz que quando a fibra de algodão tem um grau de resistência situado no escalão A então o escalão de cálcio dessa mesma fibra pertence a um escalão igualmente baixo, o escalão B. Esta regra tem uma cobertura de 25.04% e uma confiança de 85%.

Na tabela seguinte é elaborada uma síntese dos resultados obtidos, em conjunto com a análise da tabela 17.



Escalões de Ca	Escalões				Origem
	UI	ST	RD	+b	
A	Elevado	C	C	C	Zimbabwe
B	Elevado, Muito Elevado	A, B	B, D	A, B, C	Brasil, Camarões, Uganda
C	Elevado, Muito Elevado	C, D	B	B, C	Rússia, Madagáscar
D	Muito Elevado	C, D	A	A, D	Marrocos, Egipto

**Tabela 24** – Matriz escalões de Ca vs. propriedades físicas relevantes.

### 5.5.3 - K (Potássio)

Tal como no estudo das propriedades químicas anteriores, para o estudo da propriedade K (potássio) o primeiro passo consistiu na verificação das propriedades físicas relevantes para esta nova variável de *output* através da aplicação da *stream importancia.str* (figura 13). Como resultado final da aplicação desta *stream* obtiveram-se como propriedades físicas relevantes as propriedades, SL50 (Comprimento de fibra), UI (índice de uniformidade) e RD (reflectância).

De seguida e através da aplicação da *stream* da figura 14 (*esc\_total.str*) procedeu-se à criação de uma nova variável (*esc\_SL50*) mais a junção das variáveis *esc\_RD* e *esc\_UI* criadas anteriormente, controlando o nível de ocorrências por escalão gerado através de um gráfico de distribuição e obtiveram-se duas novas tabelas, de nome *tr\_total\_esc\_K* e *val\_total\_esc\_K*.

Os escalões criados referentes às novas variáveis podem ser analisados através das seguintes tabelas 20, 22 e 25.

Escalões	SL50
A	< 13.66
B	[ 13.66;14.61 [
C	[ 14.61;18.10 [
D	>= 18.10

**Tabela 25** – Escalões de SL50 gerados.



### 5.5.3.1 - RESULTADOS OBTIDOS

Aplicou-se a *stream analise\_relacoes.str* (figura 15), obtendo-se regras e posteriores resultados através da visualização de gráficos *Web*. Estes resultados estão demonstrados em anexo IV.

Para demonstrar o grau de confiança das regras geradas, vamos tomar em consideração a primeira regra gerada com o algoritmo GRI:

**esc\_K == A <= esc\_UI == Muito Elevado & esc\_RD == A (238:18.45%, 0.74)**

Esta regra diz-nos que uma determinada fibra que possui um índice de uniformidade muito elevado e um grau de reflectancia muito baixo, isto é, com valores dentro do escalão A, então essa fibra vai possuir níveis de potássio muito baixos (escalão A). Esta regra possui um grau de cobertura de 18.45% e uma confiança de 74%.

Na tabela seguinte (tabela 26) é elaborada uma síntese dos resultados obtidos, em conjunto com os dados obtidos na tabela 17.

Escalões de K	Escalões			Origem
	SL50	RD	UI	
A	D	A	Muito Elevado	Marrocos, Egipto
B	C	B, C	Muito Elevado	Rússia, Uganda
C	A, B	B	Elevado	Madagáscar, Brasil
D	A, B	D	Elevado	Zimbabwe, Camarões

**Tabela 26** – Matriz escalões de K vs. propriedades físicas relevantes.

### 5.5.4 - ME (Materiais Extraídos)

Tal como anteriormente, aplicou-se a *stream* da figura 13 (*importancia.str*) tendo como variável de *output* a propriedade química ME (materiais extraídos). Obtiveram-se como variáveis relevantes as seguintes cinco propriedades: +b (grau de amarelo), RD (reflectância), micronaire, PSI (resistência) e ST (resistência).



Com a aplicação da *stream* da figura 14 (*esc\_total.str*) procedeu-se à criação de uma nova variável (*esc\_PSI*), mais a junção das variáveis *esc\_+b*, *esc\_RD*, *esc\_micronaire* e *esc\_ST*, geradas anteriormente. Como resultado final obtiveram-se duas novas tabelas de nome *tr\_total\_esc\_ME* (teste) e *val\_total\_esc\_ME* (validação).

Os escalões criados referentes às novas variáveis podem ser analisados através das tabelas 18, 19, 20, 23 e 27.

Escalões	PSI
A	< 92.14
B	[ 92.14;97.78 [
C	[ 97.78;104.95 [
D	>= 104.95

**Tabela 27** – Escalões de PSI gerados.

#### 5.5.4.1 - RESULTADOS OBTIDOS

Com a aplicação da *stream analise\_relacoes.str* (figura 15), foram obtidas regras e a posterior visualização dos resultados obtidos através de gráficos *Web*. Estes resultados estão demonstrados em anexo V.

Podemos verificar o grau de confiança das regras geradas com o algoritmo GRI através a análise a uma das regras geradas, assim:

**esc\_ME == A <= esc\_ST == D (277:21.473%, 0.92)**

Uma determinada fibra que possua um índice de resistência muito elevado (escalão D) implica que tem uma percentagem de materiais extraídos muito baixa (escalão A). Esta regra possui uma cobertura de 21.47% e uma confiança de 92%.

Na tabela seguinte (tabela 28) é elaborada uma síntese dos resultados obtidos, em conjunto com a análise da tabela 17.



Escalões de ME	Escalões					Origem
	+b	RD	Micronaire	PSI	ST	
A	A, C	B	Fina, Média	D	C, D	Rússia, Uganda, Egipto
B	A, B, C	B, C, D	Média	A, B	A, B	Brasil, Zimbabwe, Madagáscar
C	-	D	Fina	A	A	Camarões
D	D	A	Fina	-	C	Marrocos

Tabela 28 – Matriz escalões de ME vs. propriedades físicas relevantes.

### 5.5.5 - Mg (Magnésio)

Aplicou-se a *stream* da figura 13 (*importancia.str*) tendo como variável de *output* a propriedade química Mg (magnésio). Obtiveram-se como relevantes somente as seguintes propriedades físicas: SL2.5 (comprimento) e PSI (resistência).

Aplicou-se a *stream* da figura 14 (*esc\_total.str*) procedendo-se à criação de uma nova variável (*esc\_SL25*) e a junção da variável *esc\_PSI*, anteriormente gerada. Como resultado final obtiveram-se duas novas tabelas de nome *tr\_total\_esc\_Mg* (teste) e *val\_total\_esc\_Mg* (validação).

Os escalões criados referentes às novas variáveis podem ser analisados nas tabelas 27 e 29.

Escalões	SL2.5
A	< 28.84
B	[ 28.84;30.03 [
C	[ 30.03;32.40 [
D	>= 32.40

Tabela 29 – Escalões de SL2.5 gerados.

#### 5.5.5.1 - RESULTADOS OBTIDOS

Com a aplicação da *stream analise\_relacoes.str* (figura 15), obtiveram-se as regras e resultados através da visualização de gráficos *web*, tal como está demonstrado em anexo VI.



Vamos analisar uma das regras obtidas com a aplicação do algoritmo GRI.

**esc\_Mg == A <= esc\_PSI == A (351:27.209%, 0.83)**

Através da análise a esta regra podemos afirmar que uma determinada fibra que possua um grau de resistência muito baixo (escala A) implica um grau de magnésio também muito baixo (escala A). Esta regra cobre uma percentagem de ocorrências de 27.21% e possui um grau de confiança de 83%.

Na tabela 30 é elaborada uma síntese dos resultados obtidos, em conjunto com os dados existentes na tabela 17.

Escalões de Mg	Escalões		Origem
	SL2.5	PSI	
A	A, B	A, B	Brasil, Camarões, Uganda
B	-	-	Zimbabwe
C	-	C	Rússia
D	D	D	Marrocos, Egito

**Tabela 30** – Matriz escalões de Mg vs. propriedades físicas relevantes.

### 5.5.6 - P (Fósforo)

Aplicou-se a *stream importancia.str* (figura 13) tendo como variável de *output* a propriedade química P (fósforo). Os campos referentes às propriedades físicas de menor importância para esta variável de *output* foram filtrados obtendo-se como variáveis relevantes as seguintes: PSI (resistência), SL50 (comprimento), EL (alongamento), RD (reflectância) e +b (grau de amarelo).

Com a aplicação da *stream* da figura 14 (*esc\_total.str*) procedeu-se à criação das uma nova variável (*esc\_EL*) e a junção das variáveis *esc\_PSI*, *esc\_SL50*, *esc\_RD* e *esc\_+b* geradas anteriormente. Como resultado final da aplicação desta *stream* obtiveram-se duas novas tabelas de nome *tr\_total\_esc\_P* (teste) e *val\_total\_esc\_P* (validação).

Os escalões criados referentes às novas variáveis podem ser analisados através das tabelas 18, 20, 25, 27 e 31.



Escalões	EL
Muito Baixo	< 5.0
Baixo	[ 5.0;5.8 ]
Médio	[ 5.9;6.7 ]
Elevado	[ 6.8;7.6 ]
Muito Elevado	> 7.6

Tabela 31 – Escalões de EL gerados.

### 5.5.6.1 - RESULTADOS OBTIDOS

Com a aplicação da *stream analise\_relacoes.str* (figura 15), obtiveram-se as regras resultantes da aplicação dos modelos, bem como os resultados da visualização de gráficos *web*. Estes resultados estão demonstrados em anexo VII.

Tomando como exemplo uma das regras geradas através da aplicação do modelo GRI, para assim verificarmos o grau de confiança das regras geradas, temos:

**esc\_P == A <= esc\_RD == D (307:23.798%, 0.56)**

Analisando esta regra verificamos que quando uma fibra possui um grau de reflectância muito elevado (escalão D) implica um nível muito baixo de potássio (escalão A). Esta regra possui uma cobertura de 23.80% e uma confiança de 56%.

Na tabela 32 é elaborada uma síntese dos resultados obtidos, tendo em conta os dados demonstrados na tabela 17.

Escalões de P						Origem
	PSI	SL50	EL	RD	+b	
<b>A</b>	A	B	-	D	-	Camarões
<b>B</b>	A, B, C	A, B, C	Médio, Elevado	B, C	A, B, C	Brasil, Zimbabwe, Madagáscar
<b>C</b>	D	C	Médio	B	C	Rússia, Uganda
<b>D</b>	D	D	Médio	A	D	Marrocos, Egípto

Tabela 32 – Matriz escalões de P vs. propriedades físicas relevantes.



## 5.6 - DISCUSSÃO DE RESULTADOS

Após a análise dos resultados obtidos através da aplicação das *streams* criadas no *Clementine*, foram elaboradas um conjunto de tabelas para determinar as possíveis relações existentes entre cada propriedade física considerada relevante, com cada uma das propriedades químicas em estudo.

Assim, analisaram-se todas as propriedades físicas em separado, apresentando-se os resultados nas tabelas 33, 34, 35, 36, 37, 38, 39, 40 e 41.

<b>+b</b>	<b>ME</b>	<b>Ca</b>	<b>P</b>	<b>%cinzas</b>
<b>A</b>	A, B	B, D	B	C, D
<b>B</b>	B	B, C	B	C, D
<b>C</b>	B	A, B, C	B, C	A, B, C
<b>D</b>	D	D	D	A, B

**Tabela 33** – Matriz +b (grau de amarelo).

Através da análise da tabela 33 e tomando em consideração as quatro variáveis físicas consideradas relevantes em relação à propriedade física grau de amarelo (+b), verificamos que as variáveis materiais extraídos (ME) e fósforo (P), são influenciadas de forma directa pela propriedade +b. Verifica-se que um aumento gradual dos escalões de +b proporciona um aumento também gradual dos escalões de ME e P. Em relação à variável %cinzas verifica-se uma relação inversa, isto é, um aumento dos escalões de +b proporciona um decréscimo gradual nas %cinzas extraídas. Em relação à propriedade química cálcio (Ca) e através da análise destes resultados, não nos é possível verificar a forma pela qual o grau de amarelo afecta o aumento e diminuição dos valores de cálcio existentes nas fibras de algodão.

<b>RD</b>	<b>ME</b>	<b>Ca</b>	<b>P</b>	<b>K</b>	<b>%cinzas</b>
<b>A</b>	D	D	D	A	A, B
<b>B</b>	A, B	B, C	B, C	B, C	B, C, D
<b>C</b>	B	A	B	B	C, D
<b>D</b>	B, C	B	A	D	A, B, C, D

**Tabela 34** – Matriz RD (reflectância).

De seguida é analisada a tabela 34, onde estão expressos os resultados obtidos em relação à propriedade física reflectância (RD). Tendo em consideração os resultados obtidos em relação às



cinco propriedades químicas em que o grau de reflectância é considerado relevante, podemos considerar que a propriedade física RD influencia de forma directa as propriedades químicas cálcio (Ca), fósforo (P) e Potássio (K). Analisando os resultados obtidos em relação às propriedades Ca e P, verificamos que um aumento gradual dos escalões de RD proporciona a diminuição de valores de Ca e P. O inverso é analisado em relação à propriedade K, verificando-se que um aumento nos escalões de RD proporciona um aumento dos escalões da respectiva propriedade química.

Em relação às propriedades químicas materiais extraídos (ME) e %cinzas e a sua possível relação com a propriedade física em análise, não se tornou possível verificar a influencia existente com ambas, isto devido à disparidade de valores obtidos e expressos na tabela.

<b>Micronaire</b>	<b>ME</b>	<b>%cinzas</b>
<b>Muito Fina</b>	-	-
<b>Fina</b>	A, C, D	A, B, C, D
<b>Média</b>	A, B	A, B, C, D
<b>Grossa</b>	-	-

**Tabela 35** – Matriz micronaire.

Tendo em consideração os resultados obtidos em relação à propriedade física micronaire, cujos valores estão apresentados na tabela 35, verificamos que as duas propriedades químicas, materiais extraídos (ME) e %cinzas, apesar de serem influenciadas pelo micronaire, já que foi considerado como relevante, não foi possível determinar qualquer relação entre o aumento do grau de micronaire das fibras e a variação das duas variáveis químicas aqui descritas

<b>PSI</b>	<b>ME</b>	<b>Mg</b>	<b>P</b>
<b>A</b>	B, C	A	A, B
<b>B</b>	B	A	B
<b>C</b>	-	C	B
<b>D</b>	A	D	C, D

**Tabela 36** – Matriz PSI (resistência).

Tendo em conta os valores obtidos em relação à propriedade física resistência (PSI) e apresentados na tabela 36, podemos verificar que o aumento gradual dos escalões de PSI proporcionam um aumento também gradual das propriedades químicas magnésio (MG) e fósforo (P). Em relação à



propriedade materiais extraídos (ME) a relação é inversa, isto é, um aumento dos valores de PSI leva a um decréscimo dos valores de ME.

<b>ST</b>	<b>ME</b>	<b>Ca</b>
<b>A</b>	B, C	B
<b>B</b>	B	B
<b>C</b>	A, D	A, C, D
<b>D</b>	D	C, D

**Tabela 37** – Matriz ST (resistência).

Através da análise da tabela 37 e devido à divergência de resultados obtidos e expressos na tabela, não é possível verificar a forma pela qual a propriedade física resistência afecta as propriedades químicas materiais extraídos (ME) e cálcio (Ca).

<b>SL2.5</b>	<b>Mg</b>	<b>K</b>
<b>A</b>	A	C, D
<b>B</b>	A	C, D
<b>C</b>	-	B
<b>D</b>	D	A

**Tabela 38** – Matriz SL2.5 (comprimento).

Em relação à propriedade física comprimento (SL2.5) e através da análise dos valores expressos na tabela 38, podemos concluir que escalões de SL2.5 e os escalões de magnésio (Mg) são directamente proporcionais. O aumento nos valores de SL2.5 proporcionam um decréscimo nos valores de potássio (K) nessas mesmas fibras.

<b>UI</b>	<b>Ca</b>	<b>K</b>
<b>Médio</b>	-	-
<b>Elevado</b>	A, B, C	C, D
<b>Muito Elevado</b>	B, C, D	A, B

**Tabela 39** – Matriz UI (índice de uniformidade).

Na tabela 39 estão expressos os resultados obtidos em relação à propriedade física índice de uniformidade (UI) e podemos verificar que para graus elevados de UI obtemos níveis baixos e



médios de cálcio (Ca) e níveis altos de potássio (K) nas fibras de algodão. No entanto para valores muito elevados de UI obtemos níveis de Ca médios e altos e níveis baixos de K.

<b>SL50</b>	<b>P</b>
<b>A</b>	<b>B</b>
<b>B</b>	<b>A, B</b>
<b>C</b>	<b>B, C</b>
<b>D</b>	<b>D</b>

**Tabela 40** – Matriz SL50 (comprimento).

Relativamente à propriedade física comprimento (SL50), os resultados estão expressos na tabela 40, onde se verifica que valores baixos de SL50 traduzem-se em valores baixos de fósforo (P) e valores elevados de SL50 correspondem a valores também altos de P.

<b>EL</b>	<b>P</b>
<b>Muito Baixo</b>	<b>-</b>
<b>Baixo</b>	<b>-</b>
<b>Médio</b>	<b>B, C</b>
<b>Elevado</b>	<b>B</b>
<b>Muito Elevado</b>	<b>-</b>

**Tabela 41** – Matriz EL (alongamento).

Por fim e em relação aos resultados obtidos respeitantes à propriedade física alongamento (EL) e expressos na tabela 41, verificamos que não é possível verificar qualquer comportamento relacional entre a propriedade física EL e a propriedade química fósforo (P).



## 5.7 – CONSIDERAÇÕES FINAIS

Para além das conclusões parciais realizadas ao longo da secção anterior, pretende-se na presente secção, com base nos pressupostos emitidos ao longo de todo o capítulo, agregar os resultados sectoriais obtidos de forma a validar a hipótese que norteou a realização desta dissertação.

A tabela 42 ilustra um cruzamento de todos os resultados obtidos mediante a fusão de dados expressos da tabela 33 até à 41.

	<b>+b</b>	<b>RD</b>	<b>SL2.5</b>	<b>SL50</b>	<b>UI</b>	<b>ST</b>	<b>PSI</b>	<b>EL</b>	<b>Micronaire</b>
<b>%Cinzas</b>	X	X							X
<b>Ca</b>	X	X	X		X	X			
<b>K</b>		X			X				
<b>ME</b>	X	X				X	X		X
<b>Mg</b>			X				X		
<b>P</b>	X	X		X			X	X	

**Tabela 42** – Influência entre as propriedades físicas e químicas

Através da análise da tabela anterior, podemos verificar que as propriedades físicas grau de amarelo (+b) e a reflectância (RD) influenciam quase a totalidade das propriedades químicas, à excepção do Potássio (K) em relação ao grau de amarelo, e do Magnésio (Mg) em relação a ambas.



# Cap. 6

## CONCLUSÕES E TRABALHO FUTURO

---

### 6.1 – CONCLUSÕES

Ao longo dos últimos anos têm-se verificado evoluções significativas relativamente à temática de descoberta de conhecimento em bases de dados. Essa evolução faz-se também sentir, quer ao nível das ferramentas que lhe são associadas, quer ao nível dos modelos que estas utilizam. Nesse sentido, com base na ferramenta *Clementine*, efectuou-se nesta dissertação um estudo tendente a analisar e avaliar o grau de confiança das regras obtidas sobre os dados em análise.

Esta análise centrou-se num tema de crucial importância para a área têxtil. Referimo-nos concretamente às relações existentes entre as propriedades físicas e químicas da fibra de algodão. Este procedimento sustentou-se na patente inexistência de informação relativa às referidas relações, convertendo-se desta forma num potencial problema da área têxtil.

Este trabalho reúne duas áreas perfeitamente distintas. Ao nível têxtil concentrou-se no estudo e compreensão de cada uma das propriedades do algodão. A vertente de sistemas de informação pretendeu mostrar a aplicação de técnicas de prospecção de dados, na análise de relações entre as propriedades físicas e químicas das fibras de algodão. Após a obtenção desse relacionamento, evoluiu-se no sentido de prever as características químicas das fibras de algodão com base nas suas características físicas.

Esta dissertação teve como objectivo principal demonstrar a viabilidade do uso de técnicas de prospecção de dados na inferência de propriedades químicas da fibra a partir das suas propriedades físicas. Nesse sentido recorreu-se à utilização dos algoritmos C5.0 e GRI que se baseiam em regras de associação e árvores de decisão. Foi assim possível obter-se modelos capazes de prever o relacionamento entre as propriedades em análise, sendo posteriormente utilizados na obtenção de fibras de qualidade superior.

A aplicação de duas técnicas de inferência de conhecimento distintas (algoritmos C5.0 e GRI) aos dados alvo permitiu observar resultados idênticos. Esta situação possibilitou a validação tanto do trabalho realizado como dos resultados obtidos.



Extrair mais e melhor conhecimento exige um trabalho significativo de selecção e preparação dos dados. Esta tarefa consumiu tempo significativo, uma vez que as fontes relativas aos dados utilizados eram bastante heterogéneas relativamente à forma.

Torna-se importante realçar todo o trabalho de preparação da base de dados que serviu de suporte a todos os testes realizados. Esse trabalho deveu-se à enorme variedade de fontes relativas aos dados físicos e à discrepância entre o número de registos de cada umas das propriedades físicas e químicas. Daí constata-se que as dificuldades encontradas estão mais relacionadas com o conteúdo dos dados analisados e não tanto com o processo de descoberta de conhecimento. Saliente-se a não existência ao longo de todo este trabalho, de problemas de carácter metodológico, uma vez que todas as tarefas efectuadas, em cada uma das fases, se encontrarem bem definidas à partida.

Comprovou-se também que as maiores dificuldades com que nos podemos confrontar no processo de descoberta de conhecimento, decorrem do facto dos dados terem sido armazenados com o objectivo de suportar processos operacionais, e não com o objectivo de descoberta de conhecimento.

Este trabalho apresenta uma componente inovadora no que concerne à apresentação de um estudo das relações entre as propriedades do algodão. Esta ideia sustenta-se na praticamente inexistência na literatura, de trabalhos relativos a esta temática, recorrendo a ferramentas de prospecção de dados.

A abordagem utilizada apresenta-se como um pequeno contributo científico no sentido de obter um melhor conhecimento acerca das técnicas de prospecção de dados, da ferramenta utilizada, bem como sobre a fibra natural. Subsequentemente verifica-se que os algodões analisados possuem características próprias relativamente à sua composição física e química. Estas características permitem efectuar uma classificação das fibras de acordo com a sua proveniência sustentando-nos na obtenção de regras com um elevado grau de confiança.

Analisando os resultados obtidos nas tabelas 33 e 42 podemos concluir que um aumento dos valores da propriedade física grau de amarelo (+b) pressupõe um aumento dos valores referentes à percentagem de material extraído (ME) e de fósforo (P), no entanto traduz-se por uma diminuição dos valores de percentagem de cinzas (% cinzas). Em relação à forma pela qual influencia a propriedade cálcio (Ca), não nos é possível retirar qualquer conclusão, visto que o aumento dos escalões de +b não provoca, nem um aumento nem a diminuição dos escalões de Ca, apenas se verifica a existência de valores de Ca em cada escalão de +b.



No que diz respeito à propriedade física reflectância (RD) e observando as tabelas 34 e 42, podemos concluir que um aumento desta propriedade traduz uma diminuição dos valores de fósforo (P) e um aumento dos valores de potássio (K) nas fibras. Em relação às restantes propriedades químicas, com as quais se verificam relações, isto é, percentagem de material extraído (ME), o cálcio (Ca) e a percentagem de cinzas (% cinzas), não se consegue prever o seu comportamento, visto não se verificar uma linearidade de valores com o aumento dos valores de reflectância (RD) nas fibras.

Seguidamente, na tabela 43, é demonstrada uma distribuição dos diferentes escalões de propriedades químicas por origem.

	Brasil	Zimbabwe	Rússia	Camarões	Madagáscar	Uganda	Marrocos	Egipto
D	1	3		3	1	1	2;4;5;6	4;5;6
C	1;3		1;4;5;6	1;2	1;3;5;6	1;4		
B	2;4;5	1;2;4;6	1;3	1;5	2;4	1;3;5		1
A	6	1;5	2	4;6		2;6	1;3	1;2;3

**Tabela 44** – Distribuição de propriedades químicas por origem

Legenda da tabela:

<b>% Cinzas</b>	→	1
<b>ME</b>	→	2
<b>K</b>	→	3
<b>P</b>	→	4
<b>Ca</b>	→	5
<b>Mg</b>	→	6

Através da análise da tabela anterior, é possível verificar que as fibras provenientes de Marrocos e Egipto, não possuem valores intermédios referentes às propriedades químicas analisadas. O espectro de valores situa-se entre níveis muito baixos (escalão A) e níveis muito elevados (escalão D).

Assim, analisando todos os resultados obtidos, pode-se indiciar que fibras provenientes de Marrocos e caracterizadas pela existência de valores muito elevados de grau de amarelo (+b) e de reflectância (RD), são caracterizadas por possuírem um grau muito elevado de materiais extraídos (ME), de cálcio (Ca) e de fósforo (P). Quando os valores referentes à reflectância são muito baixos, então os valores de potássio (K) e de percentagem de cinzas (% cinzas) são também muito baixos.



Caso as fibras sejam provenientes do Egito e elas possuam níveis muito elevados do grau de amarelo e de reflectância, traduzem-se em fibras com valores muito elevados de cálcio e fósforo. Se estas mesmas possuírem níveis muito baixos do grau de amarelo e de reflectância, então os valores de materiais extraídos, de potássio e de percentagem de cinzas, encontram-se também com valores muito baixos.

Os resultados atingidos demonstram que através do *Clementine*, recorrendo e aplicando técnicas de prospecção de dados, é possível identificar, analisar e prever relacionamentos entre as diferentes propriedades do algodão. Permitem também a inferência das propriedades químicas da fibra a partir das propriedades físicas existentes. Esta conclusão assenta nos resultados obtidos com base na utilização de diferentes modelos existentes na ferramenta, permitindo a obtenção de um conjunto de regras uniformes e com um elevado grau de confiança.



## 6.2 – TRABALHO FUTURO

Pelo facto do trabalho apresentado abranger duas áreas científicas distintas (área dos sistemas de informação, na sua vertente da descoberta de conhecimento nas bases de dados e da prospecção de dados propriamente dita e a área têxtil) potenciou as dificuldades sentidas ao longo de toda a dissertação. Este factor permite enumerar um conjunto de possibilidades de trabalho futuro que passarão pelo aprofundar de conceitos em cada um dos domínios referidos.

Os resultados obtidos nesta dissertação pretendem ser apenas um pequeno contributo para o conhecimento científico destas matérias. Não se pretende que sejam vistos como uma panaceia final sem possibilidade de evolução.

O sistema gerado, pela natureza dinâmica do ambiente onde está inserido, está sujeito a trabalho futuro. Assim, no ambiente da sua aplicação actual, torna-se necessário proceder à incorporação de novas fontes de dados referentes às propriedades químicas das fibras o que acarretará uma reformulação de algumas técnicas de prospecção de dados, bem como a aplicação de diferentes modelos existentes na ferramenta. Desta forma, será possível analisar com mais detalhe o estudo referente à inferência das propriedades químicas da fibra de algodão a partir das suas características físicas, utilizando o *Clementine*.

O repositório de dados existente, que serve de base a este trabalho, tem sido utilizado em outros estudos, recorrendo também ao *Clementine*, podendo assim ser efectuados trabalhos de integração dos estudos efectuados por forma a compreender melhor o comportamento de ambos os objectos de estudo (o *Clementine* e a fibra de algodão).



## Referências Bibliográficas

Abonyi, J. and F. Szeifert (2000). *Computacional Intelligence in Data Mining*. Veszprem, Hungary, University of Veszprem, Department of Process Engeneering.

Aguiar, A.,Efeito do branqueamento nas propriedades do algodão,A fibra de algodão,1997,nº 46,

Balasubramanian, P. (1989). Relation between maturity and micronaire, *Textile Research journal*.

Baptista, P. and M. J. Silva (2000). *Prospecção de Dados de Acesso a um Servidor de Notícias na Web*. Lisboa, Departamento de Infomática Faculdade de Ciências da Universidade de Lisboa.

Brachman, R. and T. Arnanand (1996). "The Process of Knowledge Discovery in Databases." AAAI/MIT Press37-58,

Bradow, J. M. and G. H. Davidonis (2000). "Quantitation of Fiber Quality and the Cotton Production-Processing Interface: A Physiologist's Perspective." Journal of Cotton Science 434-64,

Braga, A. P., A. P. L. F. Carvalho, et al., (2000). *Redes Neurais Artificiais: Teoria e Aplicações*. LTC, Belo Horizonte.

Clementine (1994-1998), Clementine Data Mining System: Reference Manual version 5.

Codd, E. F. (1993). *Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate.*, E.F. Codd and Associates.

Dias, S., R. Vasconcelos, et al., (2002). "Using Clementine Data Mining System in the Process of Analysis of Cotton Fiber Properties." IPMU 2002,

Elder, J. and D. Edwards (1995). Agente-Based Knowledge Discovery. AAI Spring Symposium on Informnation Gathering.

Fayyad, U., G. Piatetsky-Shapiro, et al., (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon.

Fayyad, U. M., G. Piatetsky-Shapiro, et al., (1996), Advances in Knowledge Discovery and Data Mining, AAI Press/ The MIT Press.

Ferreira, J. S. (1960). *A irregularidade dos fios têxteis: sua medição e análise*.

Frawley, W. J., G. Piatetsky-Shapiro, et al., (1991). "Knowledge Discovery in Databases: An Overview in Knowledge Discovery in Databases." AAAI/MIT Press1-30,



- Furter, R. and K. Douglas (1990). A medição do comprimento da fibra de algodão, do diâmetro e do conteúdo e de nep com os instrumentos AFIS.
- Goldberg, D. E. (1998). Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley.
- Hamby, S. D. (1965). The American Cotton Handbook - Volume one, third edition.
- Han, J. and M. Kamber (2001), Data Mining, Concepts and Techniques, Morgan Kaufmann Publishers.
- Klein, w. (1980), The Technology of short staple spinning - Manual of Textile Technology, Manchester - Textile Institute.
- Kohavic, R., G. John, et al., (1994). "MLC++: A Machine Learning Library in C++." IEEE Computer Society Press,
- Lucas, F. S. N. D. (2000). Ramas de algodão: Propriedades Físicas e Químicas. Engenharia Têxtil. Guimarães, Universidade do Minho - Tese de Mestrado.
- Mannila, H. (1996). Data Mining: Machine Learning, Statistic and Databases. Eight International Conference on Scientific and Statical Database Management.
- Mitchel, T. (1997), Machine Learning, McGraw-Hill.
- Morton, W. E. and j. W. S. Hearle (1975). Physical porperties of textile fibres.
- Muller, M. and D. Peyer (1991). HVI cotton fibre inspection use, methods and problems.
- Oliveira, R. d. (2000). O processo de extracção de conhecimento de bases de dados apoiado por agentes de software. Instituto de ciencias matemáticas e de computação. São Paulo, Universiade de São Paulo.
- Parker, J. (1982). The Advantages of Instrument Measurements os Cotton Fiber Properties, Texas Tech University, Lubbock Texas, USA.
- Piatetsky-Shapiro, G. (1991). Knowledge Discovery in Real Databases. AI Magazine.
- Queiroz, E. K. d. (1965). O algodão - novos processos de Produção, Comércio e Industria.
- Quinlan, J. R. (1993), C 4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc.
- Rezende, S. O., R. B. T. Oliveira, et al., (1998b). Visualization for Knowledge Discovery in Database. Ebecken: 81-95.
- Rocha, C. A. J. (1999). Redes Bayesianas para Extracção de Conhecimento de Bases de Dados, Considerando a Incorporação de Conhecimento de Fundo e o Tratamento de dados Incompletos. ICMC-USP. São Paulo, USP.



Schenek, A. and I. Janetzky (1990). The almeter test method - A possibility for fibre length measurements on raw cotton, *Melliand têxtil bereichte*.

Sen, A. and V. S. Jacob (1998). Industrial-Strength Data Warehousing. ACM.

SPSS, Generate maximum return on data in minimum time with Clementine,  
<http://www.spss.com/spssbi/clementine/>,

Tecuci, G. and Y. Kodratoff (1995). *Machine Learning and Knowledge Acquisition: Integrated approaches*.

Vasconcelos, R. M. (1986). *Planeamento e Cálculo de Custos numa Fiação*. Guimarães, Universidade do Minho.

Vasconcelos, R. M. (1993). *Contribuição à aplicação de técnicas de inteligência artificial na tecnologia da fiação*, Universidade do Minho - Tese de Doutoramento.

Wiederhold, G. (1996). *On the barrier and Future of Knowledge Discovery*: 7-11.



## **ANEXO I**

**(Relações entre a origem e  
as diferentes propriedades químicas)**

**Regras geradas com o modelo  
C5.0 para o campo Origem**

Rules for brasil:

Rule #1 for brasil:  
if esc\_P == B  
and esc\_Mg == A  
then -> brasil

Rule #2 for brasil:  
if esc\_P == B  
and esc\_Mg == D  
then -> brasil

Rules for camaroos:

Rule #1 for camaroos:  
if esc\_P == A  
then -> camaroos

Rules for egipto:

Rule #1 for egipto:  
if esc\_P == D  
and esc\_ME == A  
then -> egipto

Rules for madagascar:

Rule #1 for madagascar:  
if esc\_P == B  
and esc\_Mg == C  
then -> madagascar

Rules for marrocos:

Rule #1 for marrocos:  
if esc\_P == D  
and esc\_ME == [B C]  
then -> marrocos

Rule #2 for marrocos:

if esc\_P == D  
and esc\_ME == D  
then -> marrocos

Rules for russia:

Rule #1 for russia:  
if esc\_P == C  
and esc\_Mg == [B D]  
then -> russia

Rule #2 for russia:

if esc\_P == C  
and esc\_Mg == C  
then -> russia

Rules for uganda:

Rule #1 for uganda:  
if esc\_P == C  
and esc\_Mg == A  
then -> uganda

Rules for zimbabwe:

Rule #1 for zimbabwe:

if esc\_P == B  
and esc\_Mg == B  
then -> zimbabwe

Default : -> brasil

## **Regras geradas com o modelo GRI para o campo Origem**

origem == brasil <= esc\_Ca == B & esc\_ME == B (175:13.566%, 1.0)  
origem == brasil <= esc\_P == B & esc\_Ca == B (175:13.566%, 1.0)  
origem == brasil <= esc\_Mg == A & esc\_ME == B (175:13.566%, 1.0)  
origem == brasil <= esc\_Mg == A & esc\_Ca == B & esc\_ME == B (175:13.566%, 1.0)  
origem == brasil <= esc\_Mg == A & esc\_P == B (175:13.566%, 1.0)  
origem == brasil <= esc\_K == C & esc\_Ca == B (175:13.566%, 1.0)  
origem == brasil <= esc\_K == C & esc\_P == B & esc\_Ca == B (175:13.566%, 1.0)  
origem == brasil <= esc\_K == C & esc\_Mg == A (175:13.566%, 1.0)  
origem == zimbabwe <= esc\_Ca == A (174:13.488%, 1.0)  
origem == zimbabwe <= esc\_P == B & esc\_Ca == A (174:13.488%, 1.0)  
origem == zimbabwe <= esc\_Mg == B (174:13.488%, 1.0)  
origem == zimbabwe <= esc\_K == D & esc\_ME == B (174:13.488%, 1.0)  
origem == zimbabwe <= esc\_K == D & esc\_Ca == A (174:13.488%, 1.0)  
origem == zimbabwe <= esc\_K == D & esc\_P == B (174:13.488%, 1.0)  
origem == zimbabwe <= esc\_K == D & esc\_Mg == B (174:13.488%, 1.0)  
origem == russia <= esc\_Ca == C & esc\_ME == A (174:13.488%, 1.0)  
origem == russia <= esc\_P == C & esc\_Ca == C (174:13.488%, 1.0)  
origem == russia <= esc\_Mg == C & esc\_ME == A (174:13.488%, 1.0)  
origem == russia <= esc\_Mg == C & esc\_Ca == C & esc\_ME == A (174:13.488%, 1.0)  
origem == russia <= esc\_Mg == C & esc\_P == C (174:13.488%, 1.0)  
origem == russia <= esc\_K == B & esc\_Ca == C (174:13.488%, 1.0)  
origem == russia <= esc\_K == B & esc\_P == C & esc\_Ca == C (174:13.488%, 1.0)  
origem == russia <= esc\_K == B & esc\_Mg == C (174:13.488%, 1.0)

origem == camaroes <= esc\_ME == C (171:13.256%, 1.0)  
origem == camaroes <= esc\_Ca == B & esc\_ME == C (171:13.256%, 1.0)  
origem == camaroes <= esc\_P == A (171:13.256%, 1.0)  
origem == camaroes <= esc\_Mg == A & esc\_ME == C (171:13.256%, 1.0)  
origem == camaroes <= esc\_Mg == A & esc\_Ca == B & esc\_ME == C (171:13.256%, 1.0)  
origem == camaroes <= esc\_Mg == A & esc\_P == A (171:13.256%, 1.0)  
origem == camaroes <= esc\_K == D & esc\_ME == C (171:13.256%, 1.0)  
origem == camaroes <= esc\_K == D & esc\_Ca == B (171:13.256%, 1.0)  
origem == camaroes <= esc\_K == D & esc\_P == A (171:13.256%, 1.0)  
origem == camaroes <= esc\_K == D & esc\_Mg == A (171:13.256%, 1.0)  
origem == madagascar <= esc\_Ca == C & esc\_ME == B (154:11.938%, 1.0)  
origem == madagascar <= esc\_P == B & esc\_Ca == C (154:11.938%, 1.0)  
origem == madagascar <= esc\_Mg == C & esc\_ME == B (154:11.938%, 1.0)  
origem == madagascar <= esc\_Mg == C & esc\_Ca == C & esc\_ME == B (154:11.938%, 1.0)  
origem == madagascar <= esc\_Mg == C & esc\_P == B (154:11.938%, 1.0)  
origem == madagascar <= esc\_K == C & esc\_Ca == C (154:11.938%, 1.0)  
origem == madagascar <= esc\_K == C & esc\_P == B & esc\_Ca == C (154:11.938%, 1.0)

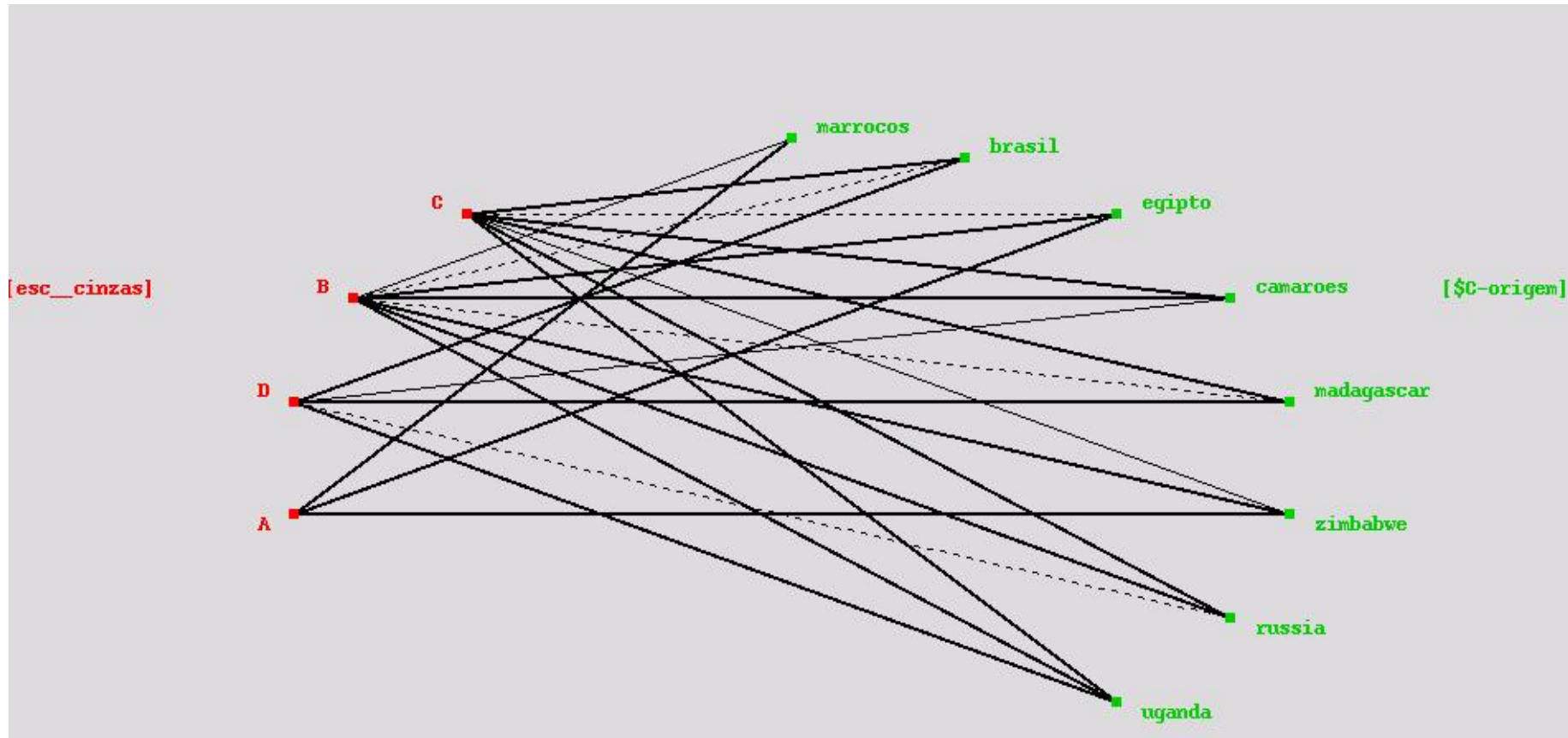
origem == madagascar <= esc\_K == C & esc\_Mg == C (154:11.938%, 1.0)  
 origem == uganda <= esc\_Ca == B & esc\_ME == A (153:11.86%, 1.0)  
 origem == uganda <= esc\_P == C & esc\_Ca == B (153:11.86%, 1.0)  
 origem == uganda <= esc\_Mg == A & esc\_ME == A (153:11.86%, 1.0)  
 origem == uganda <= esc\_Mg == A & esc\_Ca == B & esc\_ME == A (153:11.86%, 1.0)  
 origem == uganda <= esc\_Mg == A & esc\_P == C (153:11.86%, 1.0)  
 origem == uganda <= esc\_K == B & esc\_Ca == B (153:11.86%, 1.0)  
 origem == uganda <= esc\_K == B & esc\_P == C & esc\_Ca == B (153:11.86%, 1.0)  
 origem == uganda <= esc\_K == B & esc\_Mg == A (153:11.86%, 1.0)  
 origem == marrocos <= esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_Ca == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_P == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_P == D & esc\_Ca == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_Mg == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_Mg == D & esc\_Ca == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_Mg == D & esc\_P == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_K == A & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_K == A & esc\_Ca == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_K == A & esc\_P == D & esc\_ME == D (147:11.395%, 1.0)  
 origem == marrocos <= esc\_K == A & esc\_Mg == D & esc\_ME == D (147:11.395%, 1.0)

origem == egipto <= esc\_P == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_P == D & esc\_Ca == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_Mg == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_Mg == D & esc\_Ca == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_Mg == D & esc\_P == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_K == A & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_K == A & esc\_Ca == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_K == A & esc\_P == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == egipto <= esc\_K == A & esc\_Mg == D & esc\_ME == A (142:11.008%, 1.0)  
 origem == marrocos <= esc\_\_cinzas == A & esc\_ME == D (122:9.457%, 1.0)  
 origem == marrocos <= esc\_\_cinzas == A & esc\_Ca == D & esc\_ME == D (122:9.457%, 1.0)  
 origem == marrocos <= esc\_\_cinzas == A & esc\_P == D & esc\_ME == D (122:9.457%, 1.0)  
 origem == marrocos <= esc\_\_cinzas == A & esc\_Mg == D & esc\_ME == D (122:9.457%, 1.0)  
 origem == marrocos <= esc\_\_cinzas == A & esc\_K == A & esc\_ME == D (122:9.457%, 1.0)  
 origem == brasil <= esc\_\_cinzas == D & esc\_Ca == B & esc\_ME == B (111:8.605%, 1.0)  
 origem == brasil <= esc\_\_cinzas == D & esc\_P == B & esc\_Ca == B (111:8.605%, 1.0)  
 origem == brasil <= esc\_\_cinzas == D & esc\_Mg == A & esc\_ME == B (111:8.605%, 1.0)

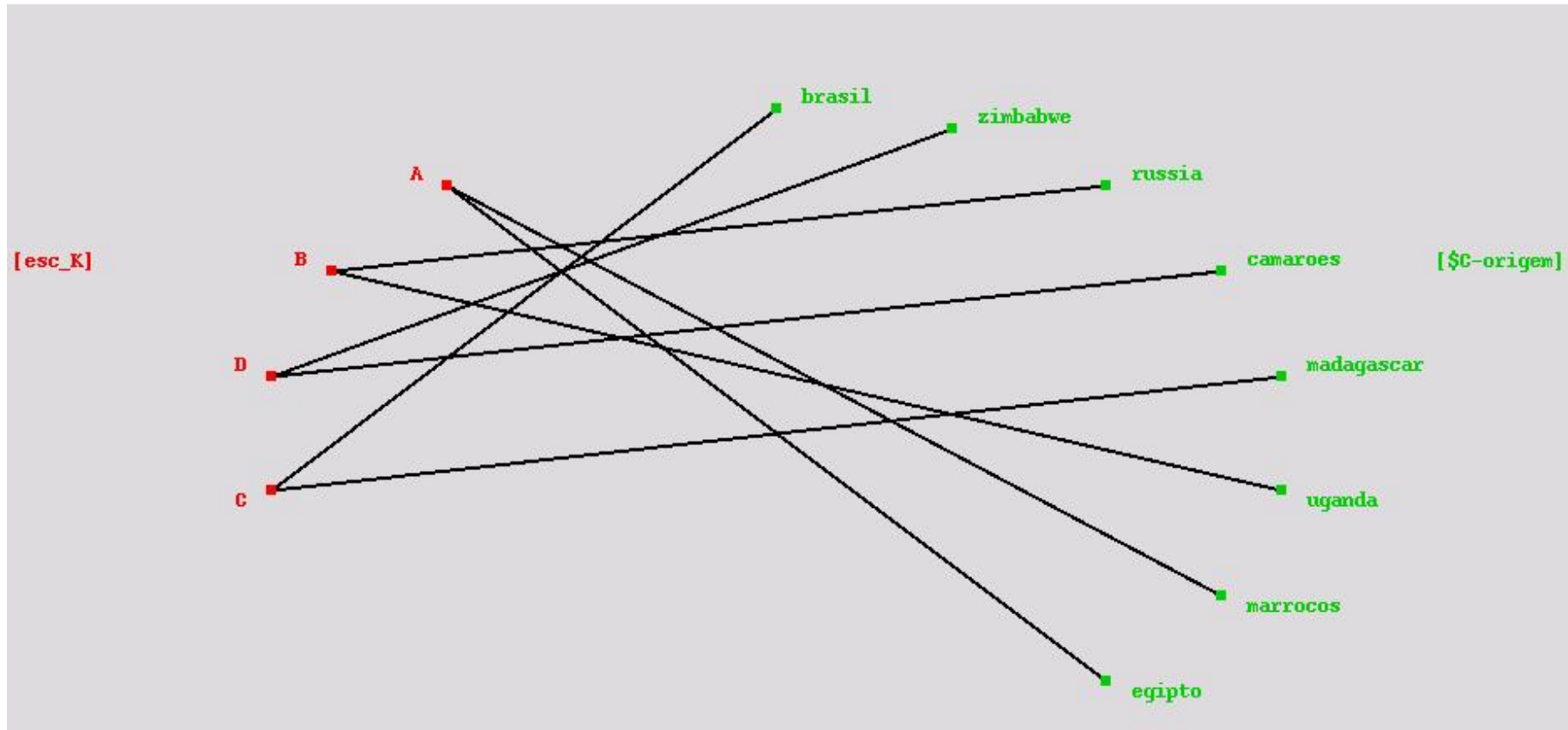
origem == brasil <= esc\_\_cinzas == D & esc\_Mg == A & esc\_P == B  
(111:8.605%, 1.0)  
origem == brasil <= esc\_\_cinzas == D & esc\_K == C & esc\_Ca == B  
(111:8.605%, 1.0)  
origem == brasil <= esc\_\_cinzas == D & esc\_K == C & esc\_Mg == A  
(111:8.605%, 1.0)  
origem == egipto <= esc\_\_cinzas == A & esc\_ME == A (99:7.674%,  
1.0)  
origem == egipto <= esc\_\_cinzas == A & esc\_Ca == D & esc\_ME == A  
(99:7.674%, 1.0)  
origem == egipto <= esc\_\_cinzas == A & esc\_P == D & esc\_ME == A  
(99:7.674%, 1.0)  
origem == egipto <= esc\_\_cinzas == A & esc\_Mg == D & esc\_ME == A  
(99:7.674%, 1.0)  
origem == egipto <= esc\_\_cinzas == A & esc\_K == A & esc\_ME == A  
(99:7.674%, 1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_ME == C (95:7.364%,  
1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_Ca == B & esc\_ME ==  
C (95:7.364%, 1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_P == A (95:7.364%,  
1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_Mg == A & esc\_ME  
== C (95:7.364%, 1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_Mg == A & esc\_P ==  
A (95:7.364%, 1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_K == D & esc\_ME ==  
C (95:7.364%, 1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_K == D & esc\_Ca == B  
(95:7.364%, 1.0)  
origem == camaroos <= esc\_\_cinzas == B & esc\_K == D & esc\_P == A  
(95:7.364%, 1.0)

origem == camaroos <= esc\_\_cinzas == B & esc\_K == D & esc\_Mg  
== A (95:7.364%, 1.0)  
origem == madagascar <= esc\_\_cinzas == D & esc\_Ca == C &  
esc\_ME == B (89:6.899%, 1.0)  
origem == madagascar <= esc\_\_cinzas == D & esc\_P == B & esc\_Ca  
== C (89:6.899%, 1.0)  
origem == madagascar <= esc\_\_cinzas == D & esc\_Mg == C &  
esc\_ME == B (89:6.899%, 1.0)  
origem == madagascar <= esc\_\_cinzas == D & esc\_Mg == C & esc\_P  
== B (89:6.899%, 1.0)  
origem == madagascar <= esc\_\_cinzas == D & esc\_K == C & esc\_Ca  
== C (89:6.899%, 1.0)

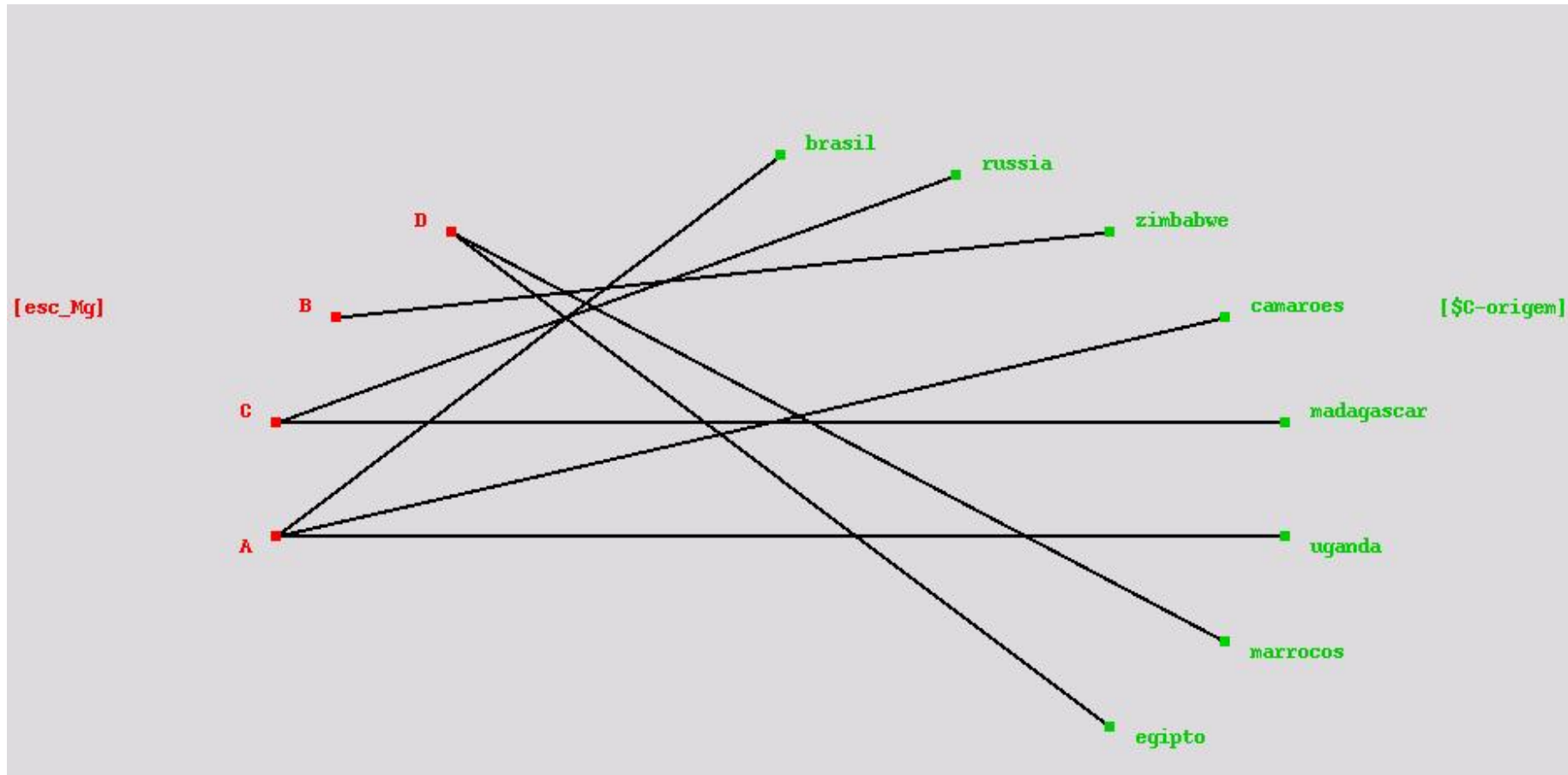
## Origem vs. Esc\_cinzas



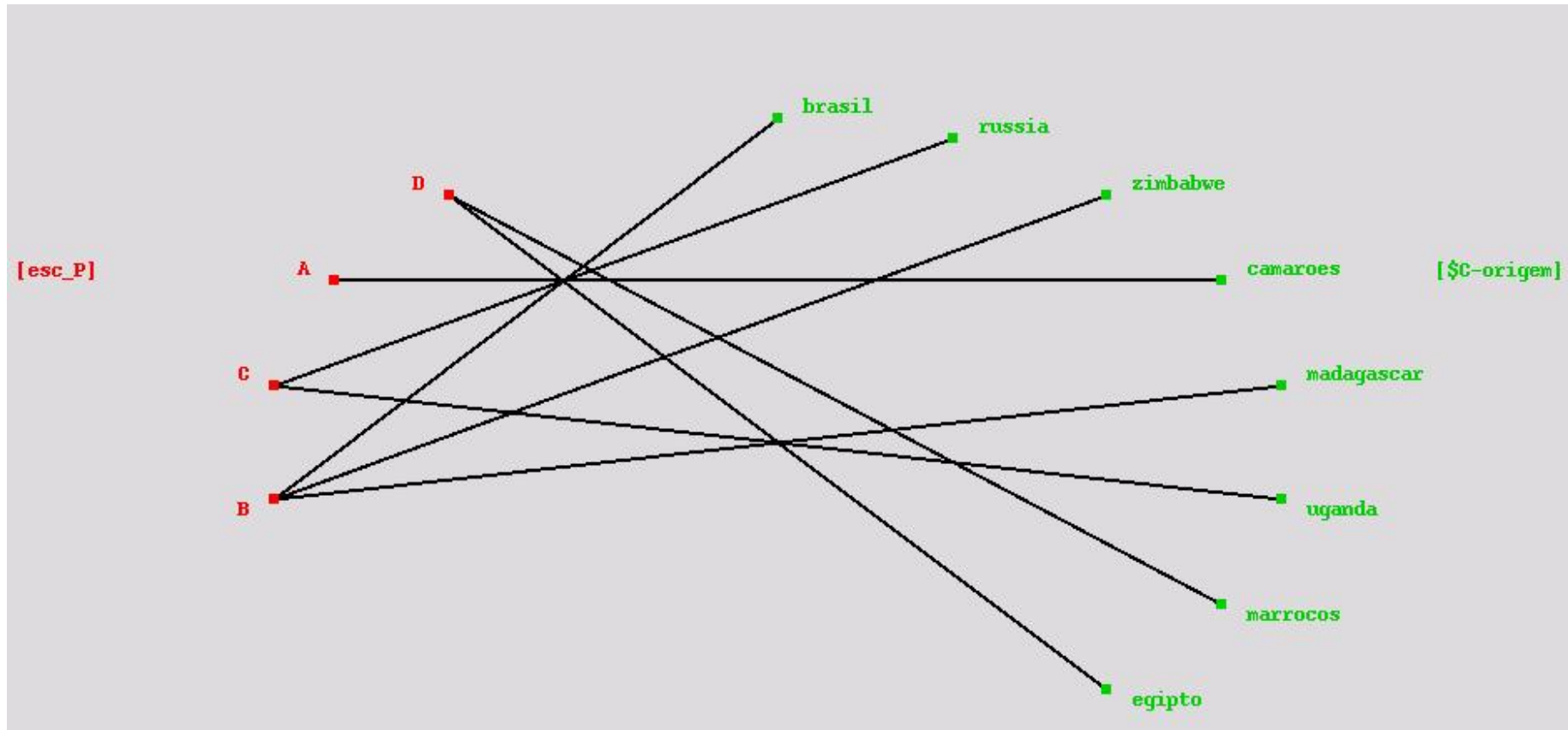
### Origem vs. K



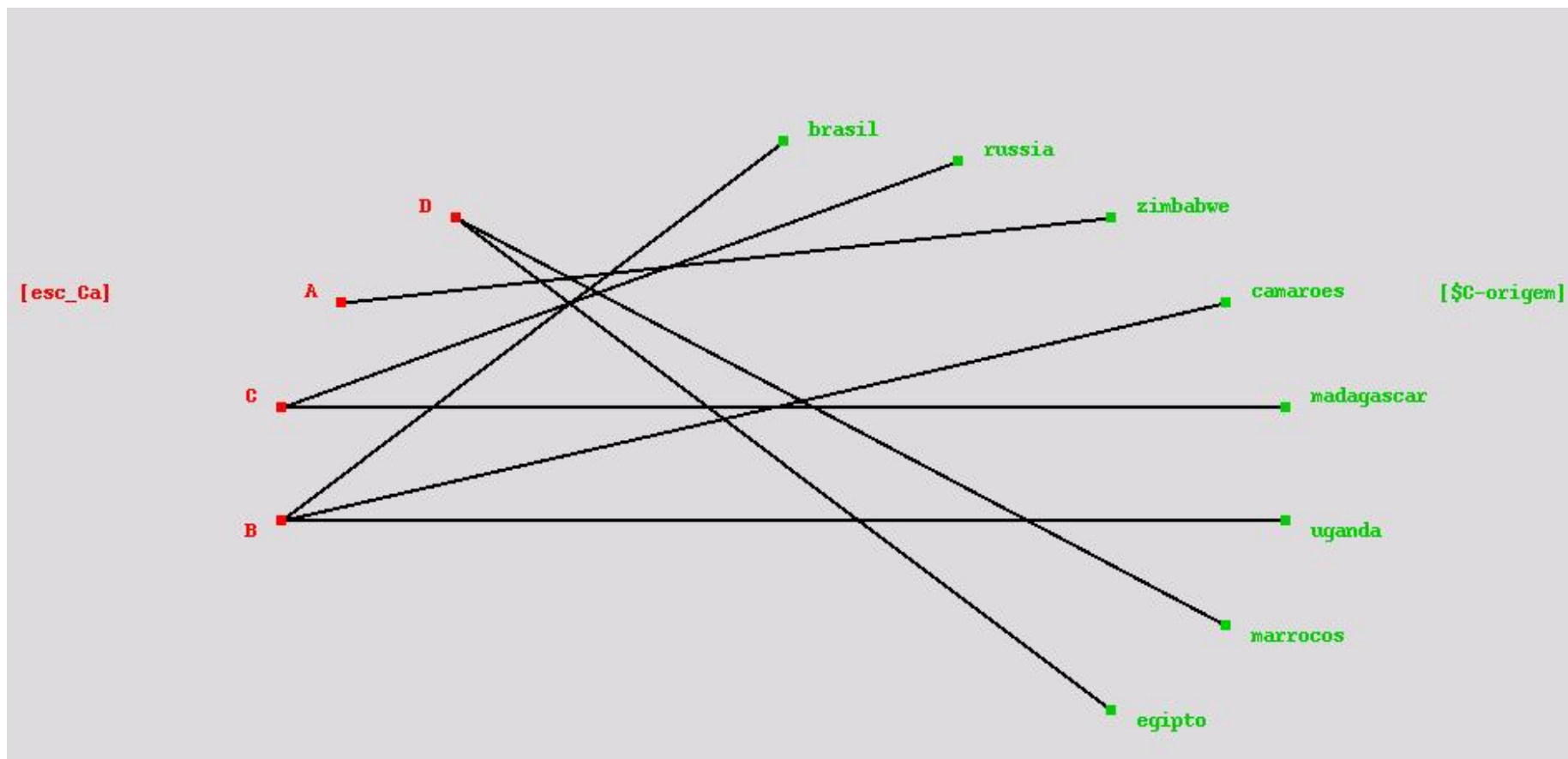
### Origem vs. Mg



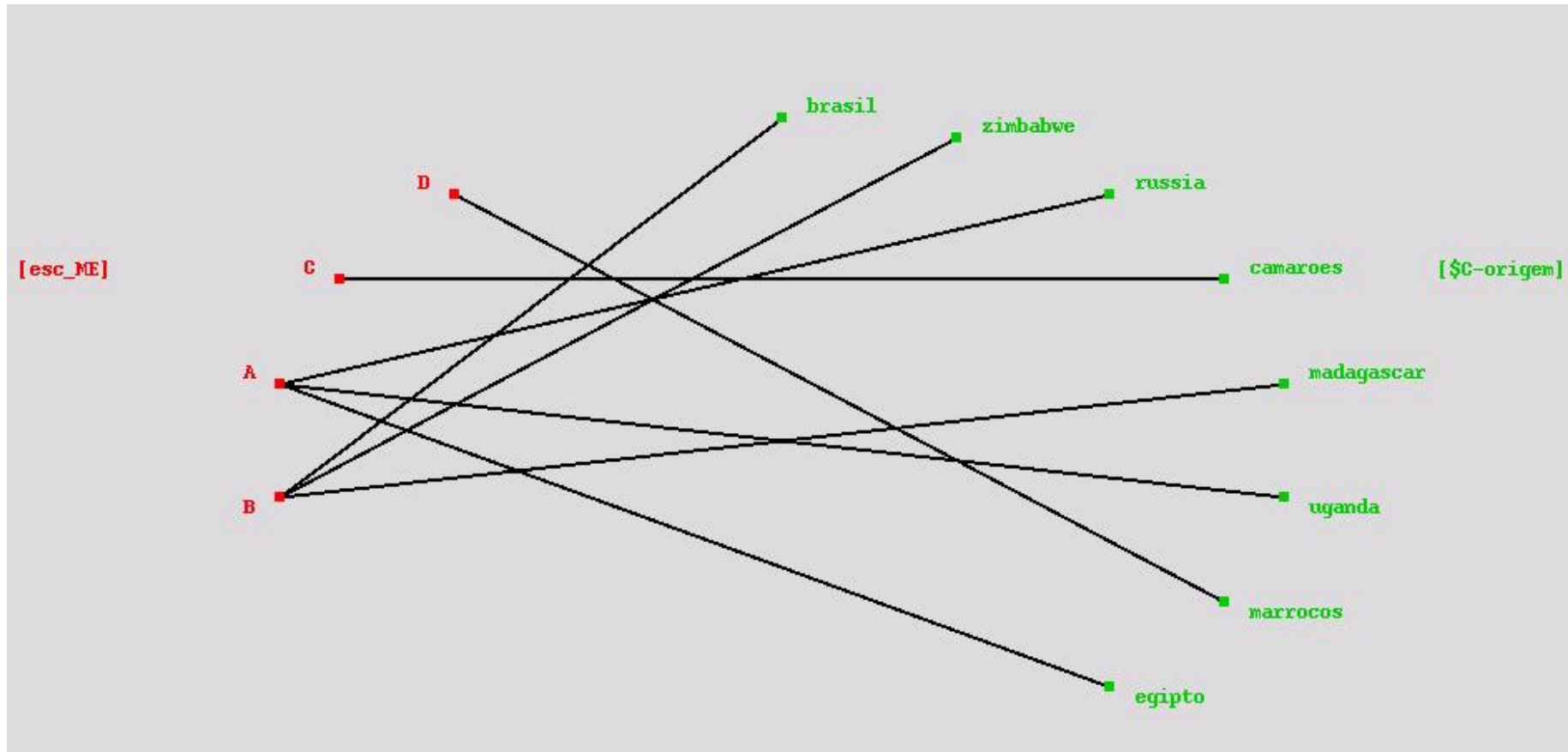
### Origem vs. P



### Origem vs. Ca



## Origem vs. ME





## **ANEXO II**

**(Resultados obtidos relativamente  
à propriedade % cinzas)**

**Regras geradas com o modelo  
C5.0 para o campo %cinzas**

Rules for A:

Rule #1 for A:

if esc\_RD == A  
and esc\_\_b == A  
then -> A

Rule #2 for A:

if esc\_RD == A  
and esc\_\_b == D  
then -> A

Rule #3 for A:

if esc\_RD == A  
and esc\_\_b == C  
and esc\_micronaire == Grossa  
then -> A

Rule #4 for A:

if esc\_RD == B  
and esc\_micronaire == Grossa  
and esc\_\_b == A  
then -> A

Rule #5 for A:

if esc\_RD == B  
and esc\_micronaire == Grossa  
and esc\_\_b == C

then -> A

Rule #6 for A:

if esc\_RD == B  
and esc\_micronaire == Grossa  
and esc\_\_b == D  
then -> A

Rule #7 for A:

if esc\_RD == C  
and esc\_micronaire == Grossa  
then -> A

Rule #8 for A:

if esc\_RD == C  
and esc\_micronaire == Media  
and esc\_\_b == D  
then -> A

Rule #9 for A:

if esc\_RD == D  
and esc\_micronaire == Grossa  
then -> A

Rules for B:

Rule #1 for B:

if esc\_RD == A  
and esc\_\_b == C  
and esc\_micronaire == Media  
then -> B

Rule #2 for B:

if esc\_RD == B  
and esc\_micronaire == Muito Fina  
then -> B

Rule #3 for B:

if esc\_RD == B  
and esc\_micronaire == Fina  
and esc\_\_b == C  
then -> B

Rule #4 for B:

if esc\_RD == B  
and esc\_micronaire == Fina  
and esc\_\_b == D  
then -> B

Rule #5 for B:

if esc\_RD == B  
and esc\_micronaire == Grossa  
and esc\_\_b == B  
then -> B

Rule #6 for B:

if esc\_RD == B  
and esc\_micronaire == Media  
and esc\_\_b == C  
then -> B

Rule #7 for B:  
if esc\_RD == C  
and esc\_micronaire == Muito Fina  
then -> B

Rule #8 for B:  
if esc\_RD == C  
and esc\_micronaire == Media  
and esc\_b == C  
then -> B

Rule #9 for B:  
if esc\_RD == D  
and esc\_micronaire == Muito Fina  
then -> B

Rule #10 for B:  
if esc\_RD == D  
and esc\_micronaire == Fina  
and esc\_b == C  
then -> B

Rule #11 for B:  
if esc\_RD == D  
and esc\_micronaire == Fina  
and esc\_b == D  
then -> B

Rule #12 for B:  
if esc\_RD == D

and esc\_micronaire == Media  
and esc\_b == D  
then -> B

Rules for C:

Rule #1 for C:  
if esc\_RD == A  
and esc\_b == C  
and esc\_micronaire == Fina  
then -> C

Rule #2 for C:  
if esc\_RD == A  
and esc\_b == C  
and esc\_micronaire == Muito Fina  
then -> C

Rule #3 for C:  
if esc\_RD == B  
and esc\_micronaire == Fina  
and esc\_b == A  
then -> C

Rule #4 for C:  
if esc\_RD == B  
and esc\_micronaire == Fina  
and esc\_b == B  
then -> C

Rule #5 for C:

if esc\_RD == B  
and esc\_micronaire == Media  
and esc\_b == D  
then -> C

Rule #6 for C:  
if esc\_RD == D  
and esc\_micronaire == Fina  
and esc\_b == A  
then -> C

Rule #7 for C:  
if esc\_RD == D  
and esc\_micronaire == Fina  
and esc\_b == B  
then -> C

Rules for D:

Rule #1 for D:  
if esc\_RD == A  
and esc\_b == B  
then -> D

Rule #2 for D:  
if esc\_RD == B  
and esc\_micronaire == Media  
and esc\_b == A  
then -> D

Rule #3 for D:

if esc\_RD == B  
and esc\_micronaire == Media  
and esc\_\_b == B  
then -> D

Rule #4 for D:  
if esc\_RD == C  
and esc\_micronaire == Fina  
then -> D

Rule #5 for D:  
if esc\_RD == C  
and esc\_micronaire == Media  
and esc\_\_b == A  
then -> D

Rule #6 for D:  
if esc\_RD == C  
and esc\_micronaire == Media  
and esc\_\_b == B  
then -> D

Rule #7 for D:  
if esc\_RD == D  
and esc\_micronaire == Media  
and esc\_\_b == A  
then -> D

Rule #8 for D:  
if esc\_RD == D

and esc\_micronaire == Media  
and esc\_\_b == B  
then -> D

Rule #9 for D:  
if esc\_RD == D  
and esc\_micronaire == Media  
and esc\_\_b == C  
then -> D

Default : -> D

### **Regras geradas com o modelo GRI para o campo %cinzas**

esc\_cinzas == A <= esc\_b == D & esc\_RD == A  
(202:15.659%, 0.68)

esc\_cinzas == A <= esc\_RD == A (274:21.24%, 0.55)

esc\_cinzas == A <= esc\_b == A & esc\_micronaire == Grossa  
(31:2.403%, 0.71)

esc\_cinzas == A <= esc\_b == A & esc\_RD == C &  
esc\_micronaire == Grossa (19:1.473%, 0.79)

esc\_cinzas == B <= esc\_b == D & esc\_RD == D (68:5.271%,  
0.62)

esc\_cinzas == B <= esc\_RD == D & esc\_micronaire == Muito  
Fina (46:3.566%, 0.59)

esc\_cinzas == D <= esc\_b == B & esc\_RD == D &  
esc\_micronaire == Media (30:2.326%, 0.53)

esc\_cinzas == B <= esc\_b == C & esc\_RD == D &  
esc\_micronaire == Fina (26:2.016%, 0.62)

esc\_cinzas == D <= esc\_b == C & esc\_RD == C &  
esc\_micronaire == Fina (17:1.318%, 0.53)

esc\_cinzas == B <= esc\_b == B & esc\_micronaire == Muito  
Fina (16:1.24%, 0.62)

esc\_cinzas == C <= esc\_b == D & esc\_RD == B &  
esc\_micronaire == Media (15:1.163%, 0.53)

esc\_cinzas == A <= esc\_b == A & esc\_RD == B &  
esc\_micronaire == Grossa (8:0.62%, 0.62)

esc\_cinzas == B <= esc\_b == B & esc\_RD == B &  
esc\_micronaire == Muito Fina (11:0.853%, 0.64)

esc\_cinzas == B <= esc\_RD == B & esc\_micronaire == Muito  
Fina (20:1.55%, 0.55)

esc\_cinzas == B <= esc\_b == D & esc\_RD == B &  
esc\_micronaire == Fina (7:0.543%, 0.71)

esc\_cinzas == B <= esc\_b == C & esc\_RD == C & esc\_micronaire ==  
Grossa (2:0.155%, 1.0)

esc\_cinzas == B <= esc\_b == B & esc\_RD == B & esc\_micronaire ==  
Grossa (4:0.31%, 0.75)

esc\_cinzas == A <= esc\_b == C & esc\_RD == A & esc\_micronaire ==  
Grossa (1:0.078%, 1.0)

esc\_cinzas == A <= esc\_b == B & esc\_RD == C & esc\_micronaire ==  
Grossa (1:0.078%, 1.0)

esc\_cinzas == C <= esc\_b == C & esc\_RD == A & esc\_micronaire ==  
Fina (3:0.233%, 0.67)

esc\_cinzas == B <= esc\_RD == C & esc\_micronaire == Muito Fina  
(5:0.388%, 0.6)

esc\_cinzas == B <= esc\_b == B & esc\_micronaire == Grossa (5:0.388%,  
0.6)

esc\_cinzas == B <= esc\_b == B & esc\_RD == D & esc\_micronaire ==  
Muito Fina (3:0.233%, 0.67)

esc\_cinzas == B <= esc\_b == A & esc\_micronaire == Muito Fina  
(3:0.233%, 0.67)

esc\_cinzas == B <= esc\_b == A & esc\_RD == C & esc\_micronaire ==  
Muito Fina (3:0.233%, 0.67)

esc\_cinzas == A <= esc\_b == C & esc\_RD == D & esc\_micronaire ==  
Grossa (4:0.31%, 0.5)

esc\_cinzas == A <= esc\_b == A & esc\_RD == D & esc\_micronaire ==  
Grossa (4:0.31%, 0.5)

esc\_cinzas == C <= esc\_b == B & esc\_RD == A & esc\_micronaire ==  
Fina (4:0.31%, 0.5)

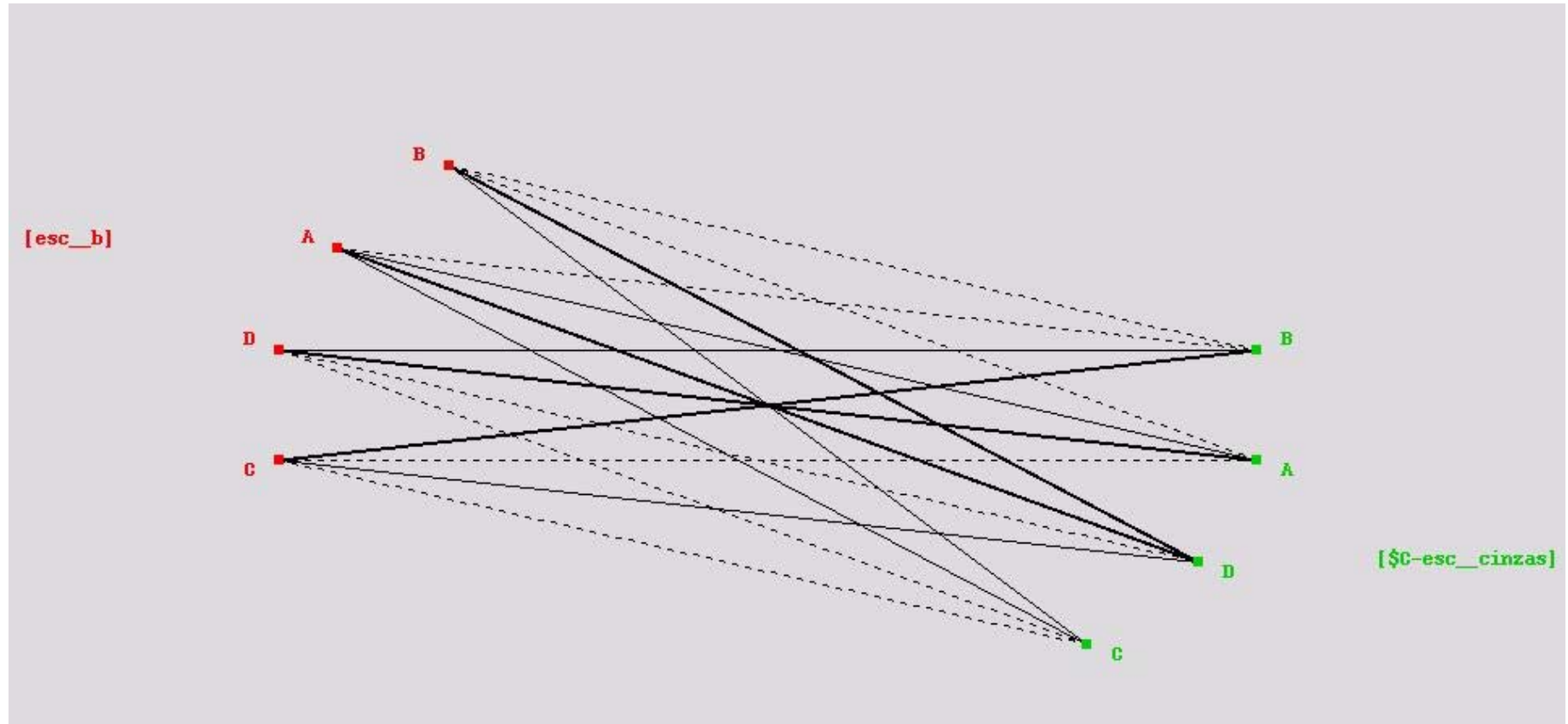
esc\_cinzas == B <= esc\_RD == D & esc\_micronaire == Grossa (8:0.62%,  
0.5)

esc\_cinzas == A <= esc\_b == C & esc\_RD == B & esc\_micronaire ==  
Grossa (2:0.155%, 0.5)

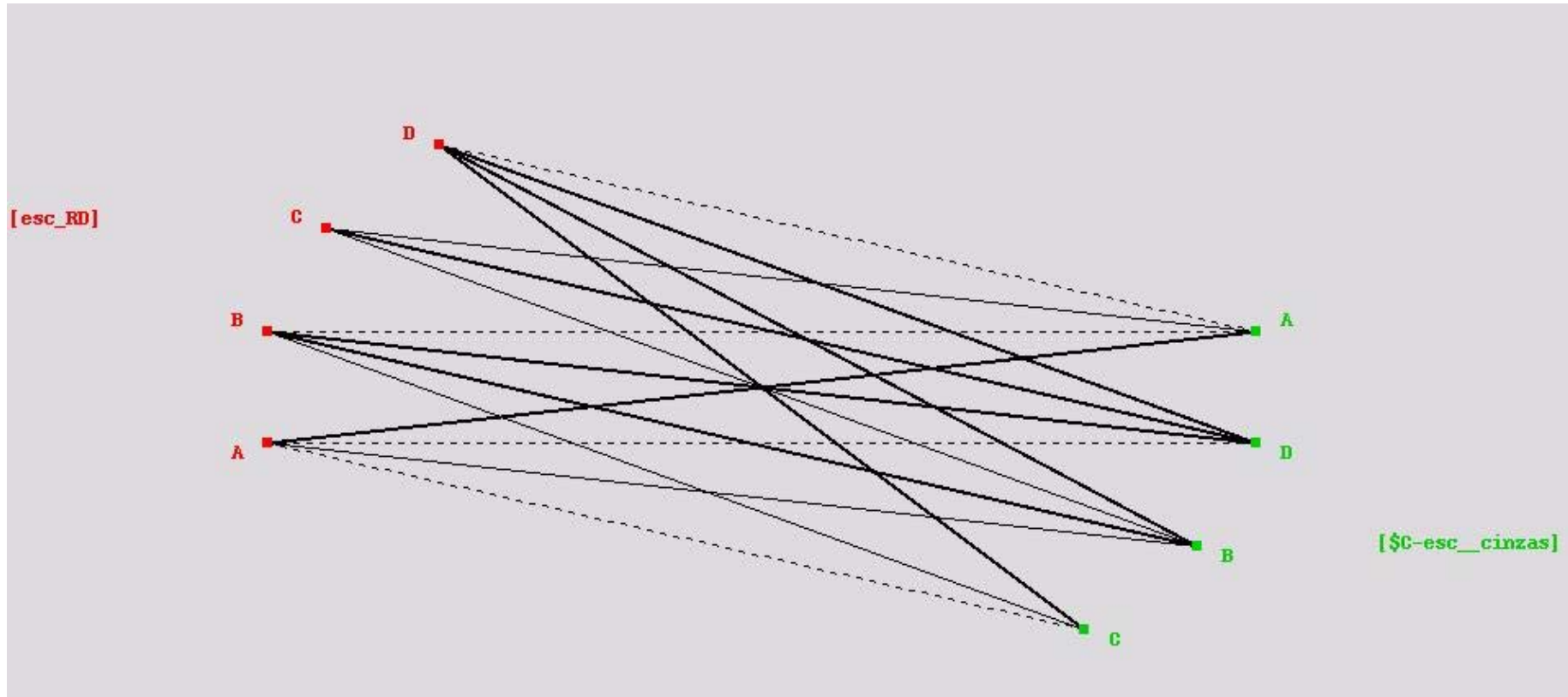
esc\_cinzas == C <= esc\_b == C & esc\_RD == B & esc\_micronaire ==  
Grossa (2:0.155%, 0.5)

esc\_\_cinzas == C <= esc\_\_b == B & esc\_RD == C &  
esc\_micronaire == Muito Fina (2:0.155%, 0.5)  
esc\_\_cinzas == D <= esc\_\_b == C & esc\_RD == A &  
esc\_micronaire == Muito Fina (2:0.155%, 0.5)  
esc\_\_cinzas == B <= esc\_\_b == D & esc\_RD == C &  
esc\_micronaire == Fina (4:0.31%, 0.5)  
esc\_\_cinzas == B <= esc\_\_b == A & esc\_RD == D &  
esc\_micronaire == Grossa (4:0.31%, 0.5)

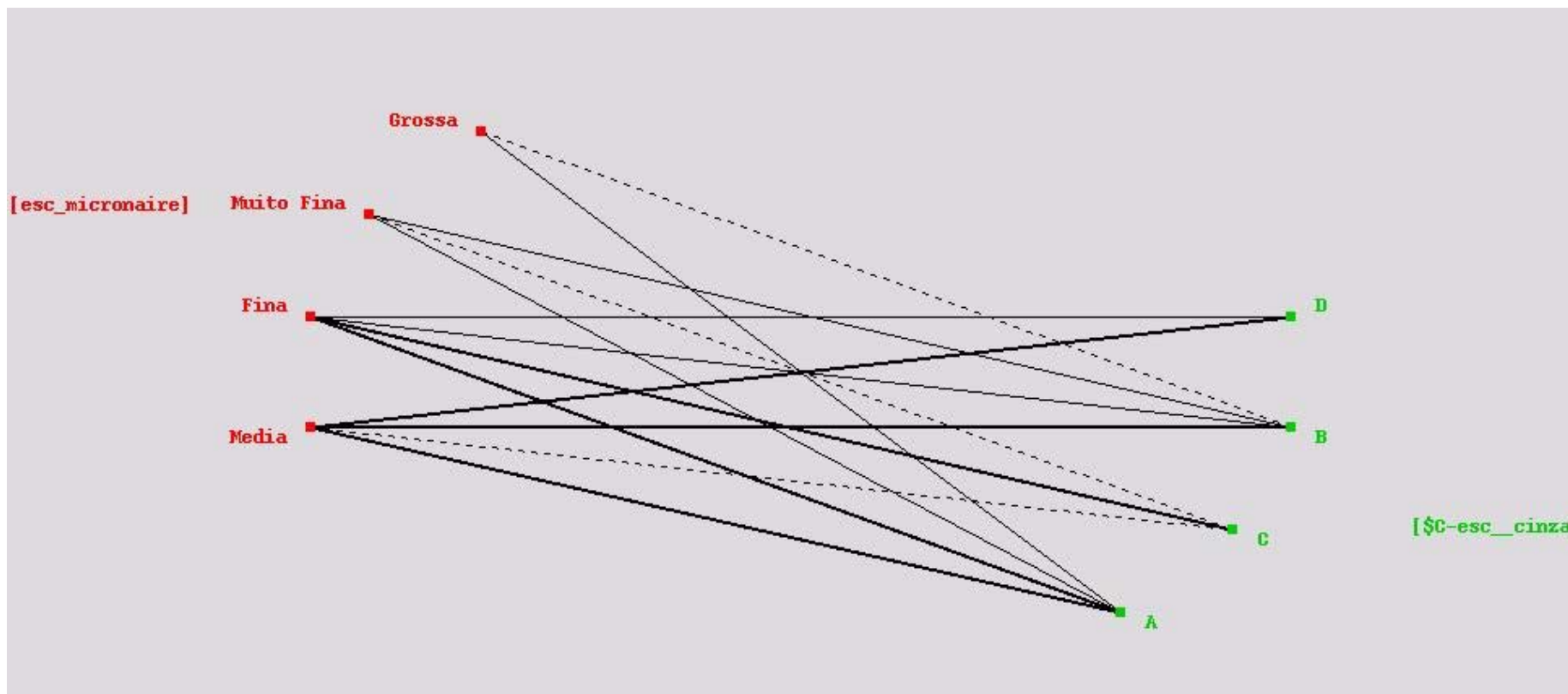
Esc\_%cinzas vs. Esc\_+b



Esc\_%cinzas vs. Esc\_RD



### Esc\_%cinzas vs. Esc\_Micronaire





## **ANEXO III**

**(Resultados obtidos relativamente  
à propriedade Ca)**

**Regras geradas com o modelo  
C5.0 para o campo Ca**

Rules for A:  
Rule #1 for A:  
if esc\_ST == A  
and esc\_UI == Muito Elevado  
and esc\_RD == C  
then -> A

Rule #2 for A:  
if esc\_ST == B  
and esc\_RD == B  
and esc\_b == C  
and esc\_UI == Elevado  
then -> A

Rule #3 for A:  
if esc\_ST == B  
and esc\_RD == B  
and esc\_b == C  
and esc\_UI == Medio  
then -> A

Rule #4 for A:  
if esc\_ST == B  
and esc\_RD == C  
and esc\_b == D  
then -> A

Rule #5 for A:  
if esc\_ST == B  
and esc\_RD == C  
and esc\_b == B  
and esc\_UI == Muito Elevado  
then -> A

Rule #6 for A:  
if esc\_ST == B  
and esc\_RD == C  
and esc\_b == C  
and esc\_UI == Muito Elevado  
then -> A

Rule #7 for A:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == D  
then -> A

Rule #8 for A:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == A  
and esc\_UI == Muito Elevado  
then -> A

Rule #9 for A:  
if esc\_ST == C  
and esc\_RD == B

and esc\_b == C  
and esc\_UI == Elevado  
then -> A

Rule #10 for A:  
if esc\_ST == C  
and esc\_RD == B  
and esc\_b == D  
and esc\_UI == Elevado  
then -> A

Rule #11 for A:  
if esc\_ST == C  
and esc\_RD == C  
and esc\_b == C  
then -> A

Rule #12 for A:  
if esc\_ST == C  
and esc\_RD == C  
and esc\_b == B  
and esc\_UI == Elevado  
then -> A

Rule #13 for A:  
if esc\_ST == C  
and esc\_RD == C  
and esc\_b == B  
and esc\_UI == Medio  
then -> A

<p>Rule #14 for A:  if esc_ST == C  and esc_RD == C  and esc__b == D  and esc_UI == Elevado  then -&gt; A</p> <p>Rule #15 for A:  if esc_ST == C  and esc_RD == C  and esc__b == D  and esc_UI == Medio  then -&gt; A</p> <p>Rule #16 for A:  if esc_ST == C  and esc_RD == D  and esc__b == C  then -&gt; A</p> <p>Rule #17 for A:  if esc_ST == C  and esc_RD == D  and esc__b == D  then -&gt; A</p> <p>Rules for B:  Rule #1 for B:  if esc_ST == A</p>	<p>and esc_UI == Elevado  then -&gt; B</p> <p>Rule #2 for B:  if esc_ST == A  and esc_UI == Medio  then -&gt; B</p> <p>Rule #3 for B:  if esc_ST == A  and esc_UI == Muito Elevado  and esc_RD == A  then -&gt; B</p> <p>Rule #4 for B:  if esc_ST == A  and esc_UI == Muito Elevado  and esc_RD == B  then -&gt; B</p> <p>Rule #5 for B:  if esc_ST == B  and esc_RD == A  then -&gt; B</p> <p>Rule #6 for B:  if esc_ST == B  and esc_RD == B  and esc__b == A  then -&gt; B</p>	<p>Rule #7 for B:  if esc_ST == B  and esc_RD == B  and esc__b == B  then -&gt; B</p> <p>Rule #8 for B:  if esc_ST == B  and esc_RD == B  and esc__b == C  and esc_UI == Muito Elevado  then -&gt; B</p> <p>Rule #9 for B:  if esc_ST == B  and esc_RD == B  and esc__b == D  and esc_UI == Elevado  then -&gt; B</p> <p>Rule #10 for B:  if esc_ST == B  and esc_RD == B  and esc__b == D  and esc_UI == Medio  then -&gt; B</p> <p>Rule #11 for B:  if esc_ST == B</p>
---	--	--

and esc\_RD == C  
and esc\_b == A  
then -> B

Rule #12 for B:  
if esc\_ST == B  
and esc\_RD == C  
and esc\_b == C  
and esc\_UI == Elevado  
then -> B

Rule #13 for B:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == B  
and esc\_UI == Elevado  
then -> B

Rule #14 for B:  
if esc\_ST == C  
and esc\_RD == B  
and esc\_b == A  
then -> B

Rule #15 for B:  
if esc\_ST == C  
and esc\_RD == B  
and esc\_b == B  
then -> B

Rule #16 for B:  
if esc\_ST == C  
and esc\_RD == C  
and esc\_b == B  
and esc\_UI == Muito Elevado  
then -> B

Rule #17 for B:  
if esc\_ST == C  
and esc\_RD == D  
and esc\_b == A  
and esc\_UI == Muito Elevado  
then -> B

Rules for C:  
Rule #1 for C:  
if esc\_ST == A  
and esc\_UI == Muito Elevado  
and esc\_RD == D  
then -> C

Rule #2 for C:  
if esc\_ST == D  
and esc\_b == B  
then -> C

Rule #3 for C:  
if esc\_ST == D  
and esc\_b == C  
then -> C

Rule #4 for C:  
if esc\_ST == D  
and esc\_b == D  
and esc\_RD == B  
then -> C

Rule #5 for C:  
if esc\_ST == D  
and esc\_b == D  
and esc\_RD == C  
then -> C

Rule #6 for C:  
if esc\_ST == B  
and esc\_RD == B  
and esc\_b == D  
and esc\_UI == Muito Elevado  
then -> C

Rule #7 for C:  
if esc\_ST == B  
and esc\_RD == C  
and esc\_b == B  
and esc\_UI == Elevado  
then -> C

Rule #8 for C:  
if esc\_ST == B  
and esc\_RD == C

and esc\_b == B  
and esc\_UI == Medio  
then -> C

Rule #9 for C:  
if esc\_ST == B  
and esc\_RD == C  
and esc\_b == C  
and esc\_UI == Medio  
then -> C

Rule #10 for C:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == C  
then -> C

Rule #11 for C:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == A  
and esc\_UI == Elevado  
then -> C

Rule #12 for C:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == A  
and esc\_UI == Medio  
then -> C

Rule #13 for C:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == B  
and esc\_UI == Medio  
then -> C

Rule #14 for C:  
if esc\_ST == B  
and esc\_RD == D  
and esc\_b == B  
and esc\_UI == Muito Elevado  
then -> C

Rule #15 for C:  
if esc\_ST == C  
and esc\_RD == A  
and esc\_b == C  
then -> C

Rule #16 for C:  
if esc\_ST == C  
and esc\_RD == B  
and esc\_b == C  
and esc\_UI == Medio  
then -> C

Rule #17 for C:  
if esc\_ST == C

and esc\_RD == B  
and esc\_b == C  
and esc\_UI == Muito Elevado  
then -> C

Rule #18 for C:  
if esc\_ST == C  
and esc\_RD == B  
and esc\_b == D  
and esc\_UI == Medio  
then -> C

Rule #19 for C:  
if esc\_ST == C  
and esc\_RD == B  
and esc\_b == D  
and esc\_UI == Muito Elevado  
then -> C

Rule #20 for C:  
if esc\_ST == C  
and esc\_RD == C  
and esc\_b == A  
then -> C

Rule #21 for C:  
if esc\_ST == C  
and esc\_RD == C  
and esc\_b == D  
and esc\_UI == Muito Elevado

```

then -> C

Rule #22 for C:
  if esc_ST == C
  and esc_RD == D
  and esc_b == B
  then -> C

Rule #23 for C:
  if esc_ST == C
  and esc_RD == D
  and esc_b == A
and esc_UI == Elevado
  then -> C

Rule #24 for C:
  if esc_ST == C
  and esc_RD == D
  and esc_b == A
  and esc_UI == Medio
  then -> C

Rules for D:
Rule #1 for D:
  if esc_ST == D
  and esc_b == A
  then -> D

Rule #2 for D:
  if esc_ST == D
  and esc_b == D
  and esc_RD == A
  then -> D

Rule #3 for D:
  if esc_ST == D
  and esc_b == D
  and esc_RD == D
  then -> D

Rule #4 for D:
  if esc_ST == C
  and esc_RD == A
  and esc_b == A
  then -> D

Rule #5 for D:
  if esc_ST == C
  and esc_RD == A
  and esc_b == B
  then -> D

Rule #6 for D:
  if esc_ST == C
  and esc_RD == A
  and esc_b == D
  then -> D

Default : -> B

```

## **Regras geradas com o modelo GRI para o campo Ca**

esc\_Ca == D <= esc\_\_b == D & esc\_RD == A (202:15.659%, 0.86)  
esc\_Ca == B <= esc\_ST == A (323:25.039%, 0.85)  
esc\_Ca == B <= esc\_UI == Elevado & esc\_ST == A (234:18.14%, 0.91)  
esc\_Ca == D <= esc\_RD == A & esc\_UI == Muito Elevado  
(238:18.45%, 0.74)  
esc\_Ca == D <= esc\_RD == A (274:21.24%, 0.69)  
esc\_Ca == B <= esc\_RD == D & esc\_ST == A (164:12.713%, 0.89)  
esc\_Ca == B <= esc\_RD == D & esc\_UI == Elevado & esc\_ST == A  
(142:11.008%, 0.92)  
esc\_Ca == B <= esc\_RD == B & esc\_ST == A (112:8.682%, 0.95)  
esc\_Ca == D <= esc\_\_b == D (322:24.961%, 0.54)  
esc\_Ca == B <= esc\_RD == B & esc\_UI == Elevado & esc\_ST == A  
(76:5.891%, 0.99)  
esc\_Ca == A <= esc\_UI == Elevado & esc\_ST == C (126:9.767%, 0.56)  
esc\_Ca == D <= esc\_\_b == A & esc\_ST == D (95:7.364%, 0.74)  
esc\_Ca == B <= esc\_\_b == D & esc\_RD == D & esc\_ST == A  
(57:4.419%, 1.0)  
esc\_Ca == B <= esc\_\_b == A & esc\_ST == A (107:8.295%, 0.87)  
esc\_Ca == B <= esc\_\_b == A & esc\_RD == B & esc\_ST == A  
(56:4.341%, 1.0)  
esc\_Ca == B <= esc\_\_b == A & esc\_RD == B & esc\_UI == Elevado  
(74:5.736%, 0.95)  
esc\_Ca == B <= esc\_\_b == A & esc\_UI == Elevado & esc\_ST == A  
(87:6.744%, 0.9)  
esc\_Ca == A <= esc\_RD == C & esc\_UI == Elevado & esc\_ST == C  
(58:4.496%, 0.71)  
esc\_Ca == D <= esc\_\_b == A & esc\_UI == Muito Elevado & esc\_ST ==  
D (89:6.899%, 0.72)  
esc\_Ca == B <= esc\_\_b == D & esc\_ST == A (61:4.729%, 0.97)  
esc\_Ca == B <= esc\_\_b == D & esc\_UI == Elevado & esc\_ST == A  
(47:3.643%, 1.0)

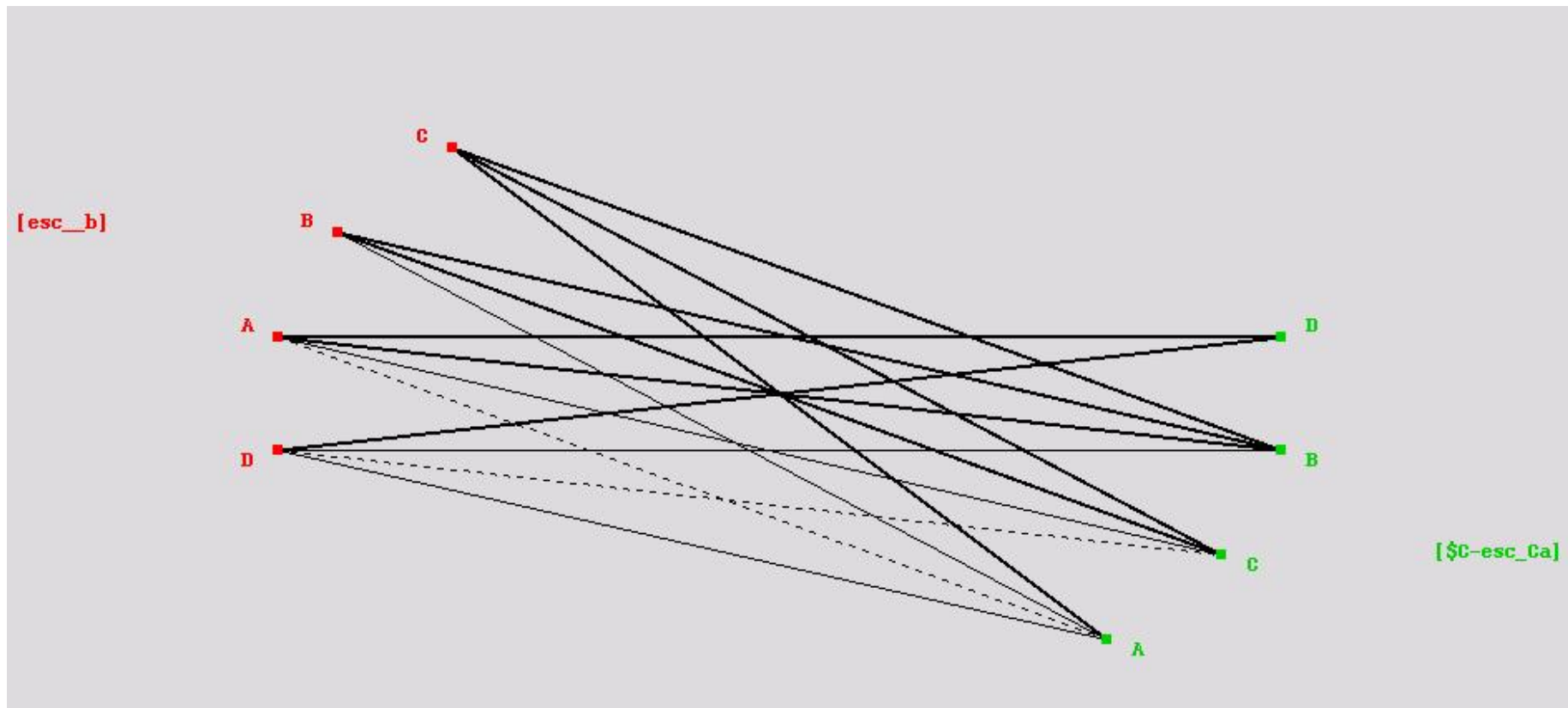
esc\_Ca == A <= esc\_\_b == C & esc\_UI == Elevado & esc\_ST == C  
(54:4.186%, 0.69)  
esc\_Ca == B <= esc\_RD == B & esc\_UI == Elevado (173:13.411%,  
0.73)  
esc\_Ca == B <= esc\_\_b == C & esc\_UI == Elevado & esc\_ST == A  
(63:4.884%, 0.9)  
esc\_Ca == B <= esc\_\_b == C & esc\_ST == A (93:7.209%, 0.82)  
esc\_Ca == B <= esc\_RD == D & esc\_UI == Elevado (242:18.76%,  
0.65)  
esc\_Ca == A <= esc\_\_b == D & esc\_RD == C & esc\_UI == Elevado  
(20:1.55%, 0.95)  
esc\_Ca == B <= esc\_\_b == C & esc\_RD == D & esc\_ST == A  
(41:3.178%, 0.98)  
esc\_Ca == B <= esc\_UI == Elevado (588:45.581%, 0.55)  
esc\_Ca == B <= esc\_\_b == A & esc\_UI == Elevado (166:12.868%,  
0.7)  
esc\_Ca == B <= esc\_\_b == D & esc\_RD == D (68:5.271%, 0.85)  
esc\_Ca == A <= esc\_\_b == D & esc\_RD == C (30:2.326%, 0.77)  
esc\_Ca == D <= esc\_\_b == A & esc\_RD == C & esc\_ST == D  
(31:2.403%, 0.87)  
esc\_Ca == C <= esc\_\_b == C & esc\_ST == D (61:4.729%, 0.7)  
esc\_Ca == B <= esc\_RD == D (307:23.798%, 0.59)  
esc\_Ca == A <= esc\_\_b == C & esc\_RD == C & esc\_ST == C  
(31:2.403%, 0.68)  
esc\_Ca == B <= esc\_\_b == B & esc\_RD == B & esc\_ST == B  
(44:3.411%, 0.89)  
esc\_Ca == C <= esc\_\_b == C & esc\_UI == Muito Elevado & esc\_ST  
== D (53:4.109%, 0.7)  
esc\_Ca == C <= esc\_\_b == C & esc\_RD == B & esc\_ST == D  
(35:2.713%, 0.8)  
esc\_Ca == C <= esc\_\_b == C & esc\_RD == A (41:3.178%, 0.76)  
esc\_Ca == C <= esc\_RD == B & esc\_ST == D (121:9.38%, 0.53)  
esc\_Ca == B <= esc\_\_b == A & esc\_RD == B (145:11.24%, 0.65)

esc\_Ca == D <= esc\_\_b == A & esc\_RD == B & esc\_ST == D  
 (50:3.876%, 0.64)  
 esc\_Ca == B <= esc\_\_b == B & esc\_ST == A (62:4.806%, 0.77)  
 esc\_Ca == B <= esc\_RD == B (414:32.093%, 0.54)  
 esc\_Ca == A <= esc\_RD == D & esc\_ST == C (48:3.721%, 0.5)  
 esc\_Ca == B <= esc\_\_b == B & esc\_UI == Elevado & esc\_ST == A  
 (37:2.868%, 0.86)  
 esc\_Ca == C <= esc\_RD == B & esc\_UI == Muito Elevado & esc\_ST  
 == D (109:8.45%, 0.52)  
 esc\_Ca == B <= esc\_RD == A & esc\_ST == B (17:1.318%, 1.0)  
 esc\_Ca == B <= esc\_\_b == C & esc\_RD == B & esc\_ST == A  
 (31:2.403%, 0.87)  
 esc\_Ca == B <= esc\_\_b == B & esc\_RD == B (121:9.38%, 0.64)  
 esc\_Ca == C <= esc\_\_b == D & esc\_RD == B & esc\_UI == Muito  
 Elevado (14:1.085%, 0.93)  
 esc\_Ca == A <= esc\_\_b == C & esc\_RD == C & esc\_UI == Muito  
 Elevado (29:2.248%, 0.55)  
 esc\_Ca == A <= esc\_RD == B & esc\_UI == Elevado & esc\_ST == C  
 (23:1.783%, 0.61)  
 esc\_Ca == A <= esc\_\_b == C & esc\_RD == D & esc\_ST == C  
 (18:1.395%, 0.67)  
 esc\_Ca == B <= esc\_RD == A & esc\_UI == Elevado & esc\_ST == B  
 (14:1.085%, 1.0)  
 esc\_Ca == C <= esc\_\_b == B & esc\_RD == C & esc\_UI == Elevado  
 (43:3.333%, 0.63)  
 esc\_Ca == D <= esc\_\_b == A & esc\_RD == B & esc\_UI == Muito  
 Elevado (62:4.806%, 0.52)  
 esc\_Ca == A <= esc\_RD == C & esc\_UI == Muito Elevado & esc\_ST  
 == A (22:1.705%, 0.59)  
 esc\_Ca == A <= esc\_UI == Medio & esc\_ST == C (14:1.085%, 0.71)  
 esc\_Ca == C <= esc\_\_b == B & esc\_ST == D (50:3.876%, 0.58)  
 esc\_Ca == C <= esc\_RD == D & esc\_ST == B (84:6.512%, 0.5)

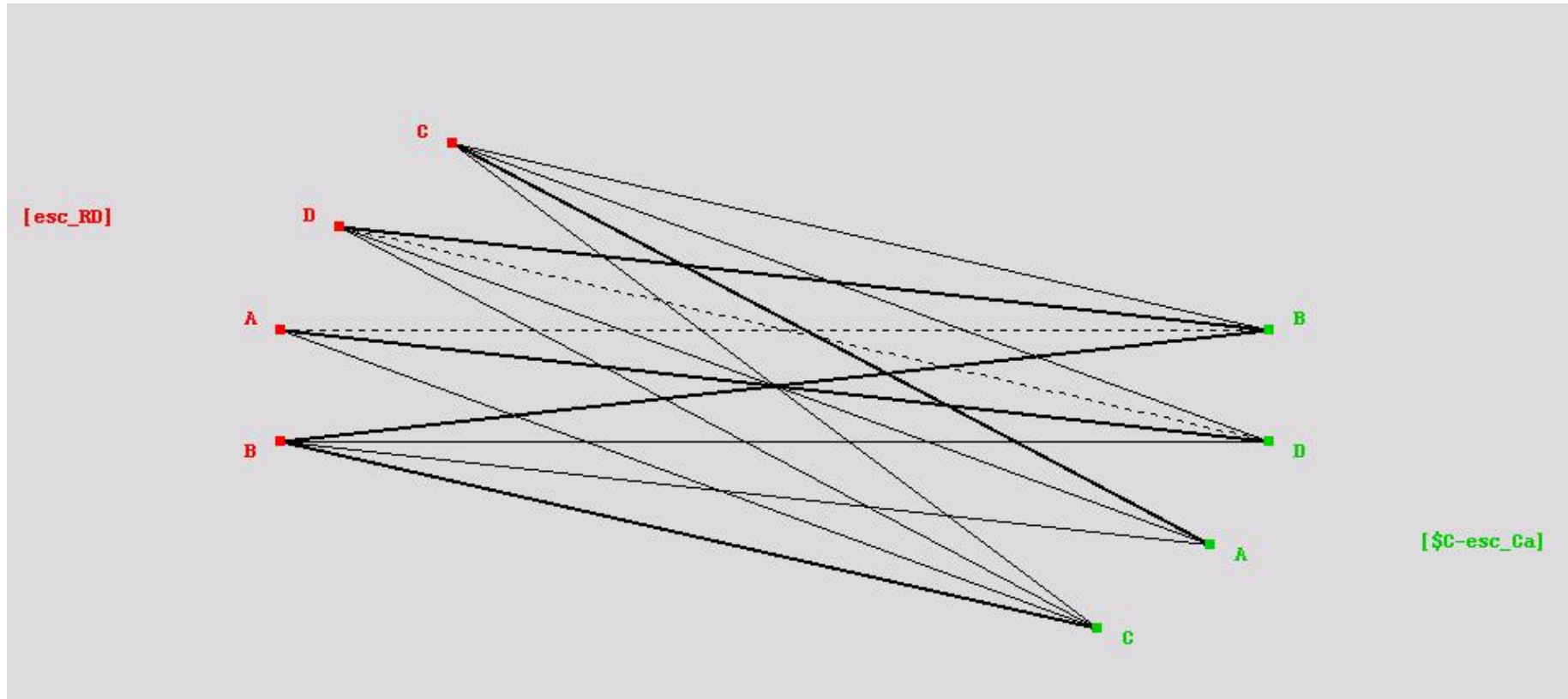
esc\_Ca == A <= esc\_RD == D & esc\_UI == Elevado & esc\_ST == C  
 (30:2.326%, 0.5)  
 esc\_Ca == B <= esc\_\_b == A (333:25.814%, 0.51)  
 esc\_Ca == A <= esc\_\_b == C & esc\_UI == Medio & esc\_ST == C  
 (5:0.388%, 1.0)  
 esc\_Ca == C <= esc\_RD == D & esc\_UI == Elevado & esc\_ST == B  
 (70:5.426%, 0.5)  
 esc\_Ca == D <= esc\_\_b == A & esc\_UI == Elevado & esc\_ST == D  
 (6:0.465%, 1.0)  
 esc\_Ca == A <= esc\_RD == D & esc\_UI == Medio & esc\_ST == C  
 (12:0.93%, 0.67)  
 esc\_Ca == C <= esc\_\_b == D & esc\_RD == B (22:1.705%, 0.68)  
 esc\_Ca == C <= esc\_\_b == B & esc\_UI == Muito Elevado & esc\_ST  
 == D (42:3.256%, 0.52)  
 esc\_Ca == C <= esc\_\_b == B & esc\_RD == D & esc\_UI == Medio  
 (5:0.388%, 1.0)  
 esc\_Ca == D <= esc\_\_b == A & esc\_RD == A & esc\_ST == D  
 (4:0.31%, 1.0)  
 esc\_Ca == C <= esc\_\_b == B & esc\_RD == B & esc\_ST == D  
 (32:2.481%, 0.53)  
 esc\_Ca == A <= esc\_\_b == C & esc\_RD == D & esc\_UI == Medio  
 (8:0.62%, 0.62)  
 esc\_Ca == D <= esc\_\_b == A & esc\_RD == D & esc\_ST == D  
 (10:0.775%, 0.7)  
 esc\_Ca == C <= esc\_\_b == A & esc\_RD == D & esc\_ST == B  
 (23:1.783%, 0.57)  
 esc\_Ca == C <= esc\_\_b == B & esc\_UI == Elevado & esc\_ST == D  
 (6:0.465%, 0.83)  
 esc\_Ca == C <= esc\_UI == Elevado & esc\_ST == D (22:1.705%, 0.55)  
 esc\_Ca == C <= esc\_\_b == B & esc\_UI == Medio & esc\_ST == C  
 (3:0.233%, 1.0)  
 esc\_Ca == A <= esc\_RD == C & esc\_UI == Medio & esc\_ST == C  
 (2:0.155%, 1.0)

esc\_Ca == D <= esc\_b == A & esc\_RD == A & esc\_UI == Muito Elevado (5:0.388%, 0.8)  
esc\_Ca == C <= esc\_b == A & esc\_UI == Elevado & esc\_ST == C (16:1.24%, 0.56)  
esc\_Ca == C <= esc\_RD == C & esc\_UI == Medio & esc\_ST == B (9:0.698%, 0.67)  
esc\_Ca == C <= esc\_b == B & esc\_UI == Medio & esc\_ST == B (5:0.388%, 0.8)  
esc\_Ca == C <= esc\_RD == D & esc\_UI == Muito Elevado & esc\_ST == A (7:0.543%, 0.71)  
esc\_Ca == C <= esc\_UI == Medio & esc\_ST == B (19:1.473%, 0.53)  
esc\_Ca == C <= esc\_RD == C & esc\_UI == Elevado & esc\_ST == A (13:1.008%, 0.54)  
esc\_Ca == C <= esc\_RD == D & esc\_UI == Medio & esc\_ST == B (6:0.465%, 0.67)  
esc\_Ca == C <= esc\_b == B & esc\_RD == A & esc\_ST == D (6:0.465%, 0.67)  
esc\_Ca == C <= esc\_RD == B & esc\_UI == Elevado & esc\_ST == D (11:0.853%, 0.55)  
esc\_Ca == D <= esc\_b == C & esc\_RD == C & esc\_ST == D (5:0.388%, 0.6)  
esc\_Ca == C <= esc\_b == B & esc\_RD == A & esc\_UI == Muito Elevado (7:0.543%, 0.57)  
esc\_Ca == C <= esc\_RD == B & esc\_UI == Medio & esc\_ST == D (1:0.078%, 1.0)  
esc\_Ca == C <= esc\_b == A & esc\_UI == Medio & esc\_ST == C (1:0.078%, 1.0)

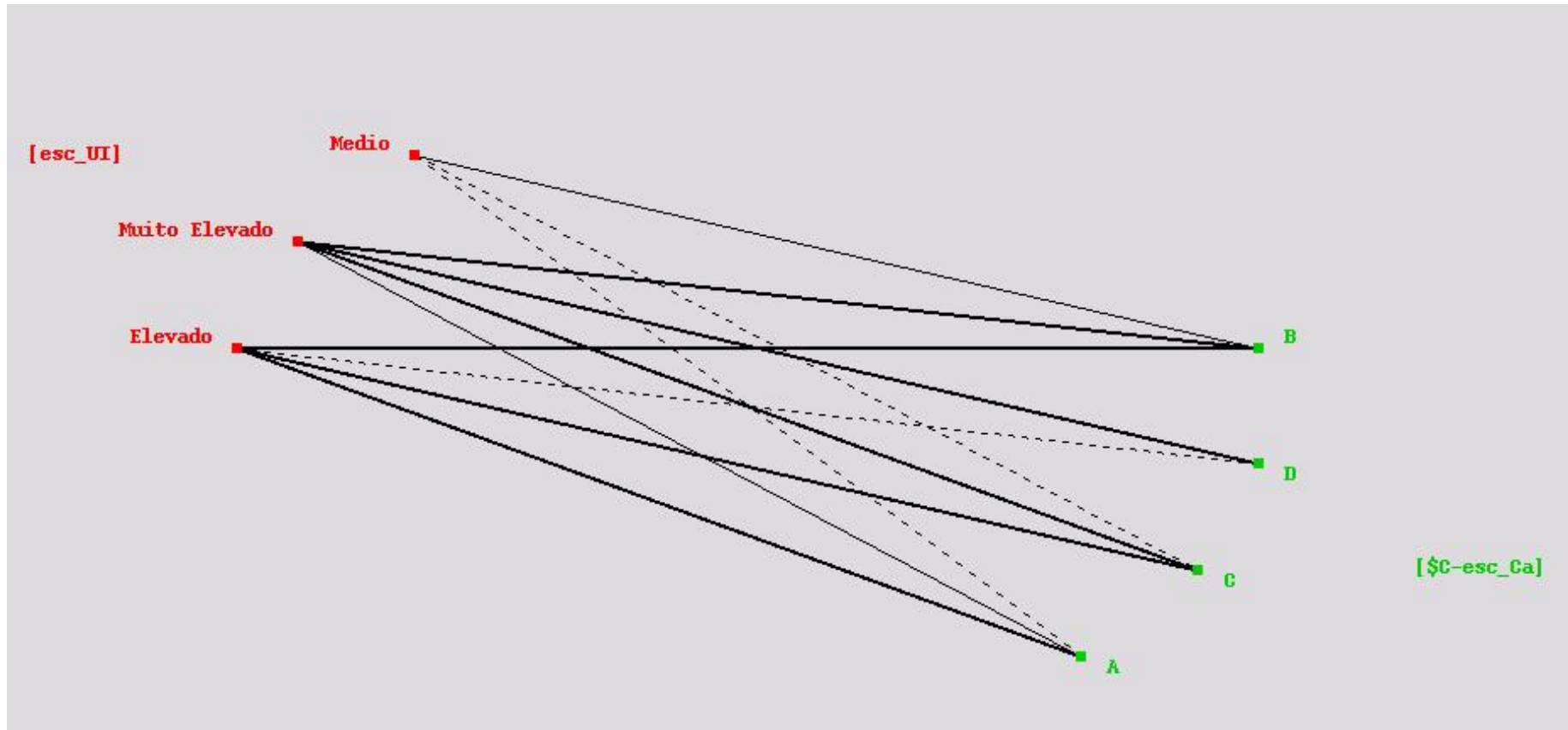
Esc\_Ca vs. Esc\_+b



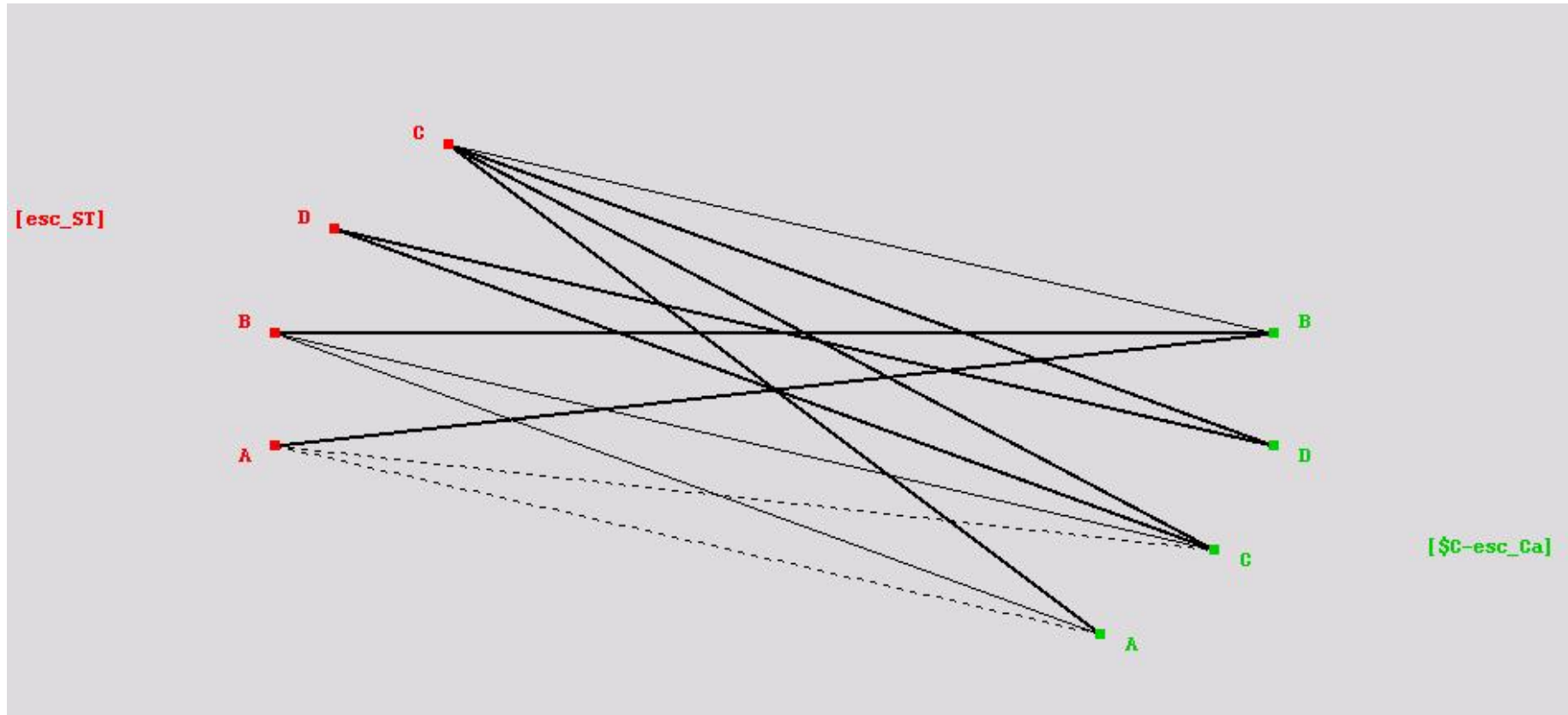
### Esc\_Ca vs. Esc\_RD



### Esc\_Ca vs. Esc\_UI



### Esc\_Ca vs. Esc\_ST





## **ANEXO IV**

**(Resultados obtidos relativamente  
à propriedade K)**

**Regras geradas com o modelo  
C5.0 para o campo K**

Rules for A:

Rule #1 for A:

if SL50 > 15  
and SL50 =< 15.5  
and esc\_RD == A  
then -> A

Rule #2 for A:

if SL50 > 14.7  
and SL50 =< 15.5  
and esc\_RD == B  
and esc\_UI == Muito  
Elevado  
then -> A

Rule #3 for A:

if SL50 > 19.3  
and esc\_RD == B  
then -> A

Rule #4 for A:

if SL50 > 15.5  
and esc\_RD == A  
and esc\_UI == Medio  
then -> A

Rule #5 for A:

if SL50 > 15.5  
and esc\_RD == A  
and esc\_UI == Muito  
Elevado  
then -> A

Rule #6 for A:

if SL50 > 15.5  
and SL50 =< 16  
and esc\_RD == A  
and esc\_UI == Elevado  
then -> A

Rule #7 for A:

if SL50 > 19.1  
and esc\_RD == D  
and esc\_SL50 == D  
then -> A

Rule #8 for A:

if SL50 > 17.8  
and SL50 =< 19.3  
and esc\_RD == C  
and esc\_SL50 == C  
then -> A

Rule #9 for A:

if SL50 > 18.3  
and SL50 =< 19.3  
and esc\_RD == C

and esc\_SL50 == D  
then -> A

Rule #10 for A:

if SL50 > 19.3  
and esc\_RD == C  
then -> A

Rules for B:

Rule #1 for B:

if SL50 =< 15.5  
and esc\_RD == B  
and esc\_UI == Muito Elevado  
then -> B

Rule #2 for B:

if SL50 > 15  
and SL50 =< 15.5  
and esc\_RD == B  
then -> B

Rule #3 for B:

if SL50 > 14.7  
and SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Elevado  
and esc\_SL50 == C  
then -> B

Rule #4 for B:

if SL50 > 15.5  
and SL50 =< 19.3  
and esc\_RD == B  
then -> B

Rule #5 for B:  
if SL50 > 16  
and esc\_RD == A  
and esc\_UI == Elevado  
then -> B

Rule #6 for B:  
if SL50 > 15.5  
and esc\_RD == D  
and esc\_SL50 == [A B]  
then -> B

Rule #7 for B:  
if SL50 > 15.5  
and SL50 =< 19.1  
and esc\_RD == D  
and esc\_SL50 == D  
then -> B

Rule #8 for B:  
if SL50 > 16.3  
and esc\_RD == D  
and esc\_SL50 == C  
and esc\_UI == Medio  
then -> B

Rule #9 for B:  
if SL50 > 16.3  
and esc\_RD == D  
and esc\_SL50 == C  
and esc\_UI == Muito Elevado  
then -> B

Rule #10 for B:  
if SL50 > 15.5  
and SL50 =< 17.5  
and esc\_RD == C  
then -> B

Rule #11 for B:  
if SL50 > 17.5  
and SL50 =< 19.3  
and esc\_RD == C  
and esc\_SL50 == D  
then -> B

Rules for C:  
Rule #1 for C:  
if SL50 =< 15.5  
and esc\_RD == A  
then -> C

Rule #2 for C:  
if SL50 =< 15.5

and esc\_RD == B  
and esc\_UI == Elevado  
then -> C

Rule #3 for C:  
if SL50 =< 15.5  
and esc\_RD == B  
and esc\_UI == Medio  
then -> C

Rule #4 for C:  
if SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Medio  
then -> C

Rule #5 for C:  
if SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Muito Elevado  
then -> C

Rule #6 for C:  
if SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Elevado  
and esc\_SL50 == D  
then -> C

Rule #7 for C:

if SL50 > 13.2  
and SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Elevado  
and esc\_SL50 == A  
then -> C

Rule #8 for C:  
if SL50 > 14.2  
and SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Elevado  
and esc\_SL50 == B  
then -> C

Rule #9 for C:  
if SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Elevado  
and esc\_SL50 == C  
then -> C

Rule #10 for C:  
if SL50 > 13  
and SL50 =< 15.5  
and esc\_RD == D  
and esc\_UI == Medio  
then -> C

Rule #11 for C:

if SL50 > 14.5  
and SL50 =< 15.5  
and esc\_RD == D  
then -> C

Rule #12 for C:  
if SL50 > 15.5  
and SL50 =< 16.3  
and esc\_RD == D  
and esc\_SL50 == C  
then -> C

Rules for D:  
Rule #1 for D:  
if SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Elevado  
and esc\_SL50 == A  
then -> D

Rule #2 for D:  
if SL50 =< 15.5  
and esc\_RD == C  
and esc\_UI == Elevado  
and esc\_SL50 == B  
then -> D

Rule #3 for D:  
if SL50 =< 15.5  
and esc\_RD == D

and esc\_UI == Elevado  
then -> D

Rule #4 for D:  
if SL50 =< 15.5  
and esc\_RD == D  
and esc\_UI == Muito Elevado  
then -> D

Rule #5 for D:  
if SL50 =< 15.5  
and esc\_RD == D  
and esc\_UI == Medio  
then -> D

Rule #6 for D:  
if SL50 > 13.2  
and SL50 =< 15.5  
and esc\_RD == D  
and esc\_UI == Medio  
then -> D

Rule #7 for D:  
if SL50 > 16.3  
and esc\_RD == D  
and esc\_SL50 == C  
and esc\_UI == Elevado  
then -> D

Rule #8 for D:

```
    if SL50 > 17.5
    and SL50 =< 19.3
    and esc_RD == C
and esc_SL50 == [A B]
    then -> D
```

```
Rule #9 for D:
    if SL50 > 17.5
    and SL50 =< 19.3
    and esc_RD == C
and esc_SL50 == C
    then -> D
```

```
Rule #10 for D:
    if SL50 > 19.1
    and SL50 =< 19.3
    and esc_RD == C
    then -> D
```

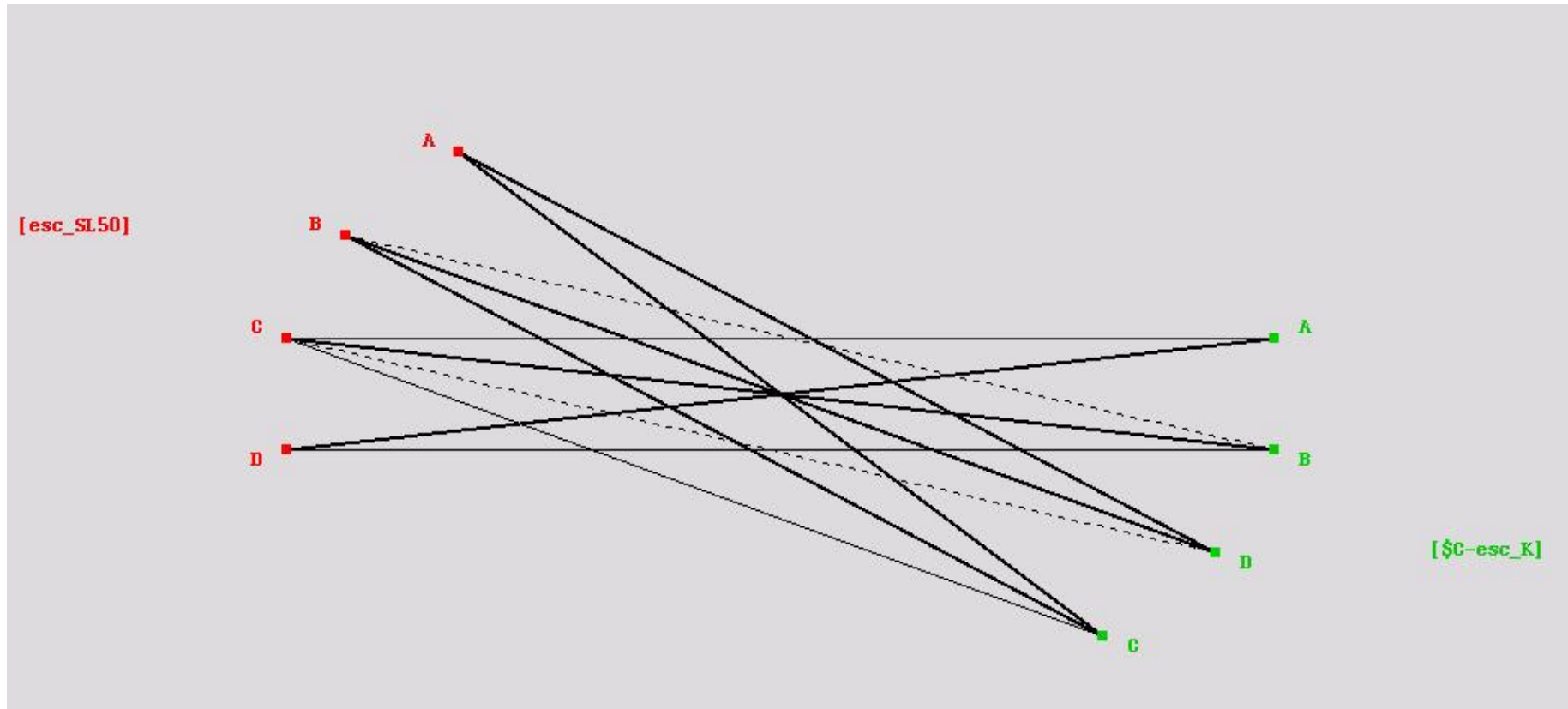
```
Default : -> A
```

### **Regras geradas com o modelo GRI para o campo K**

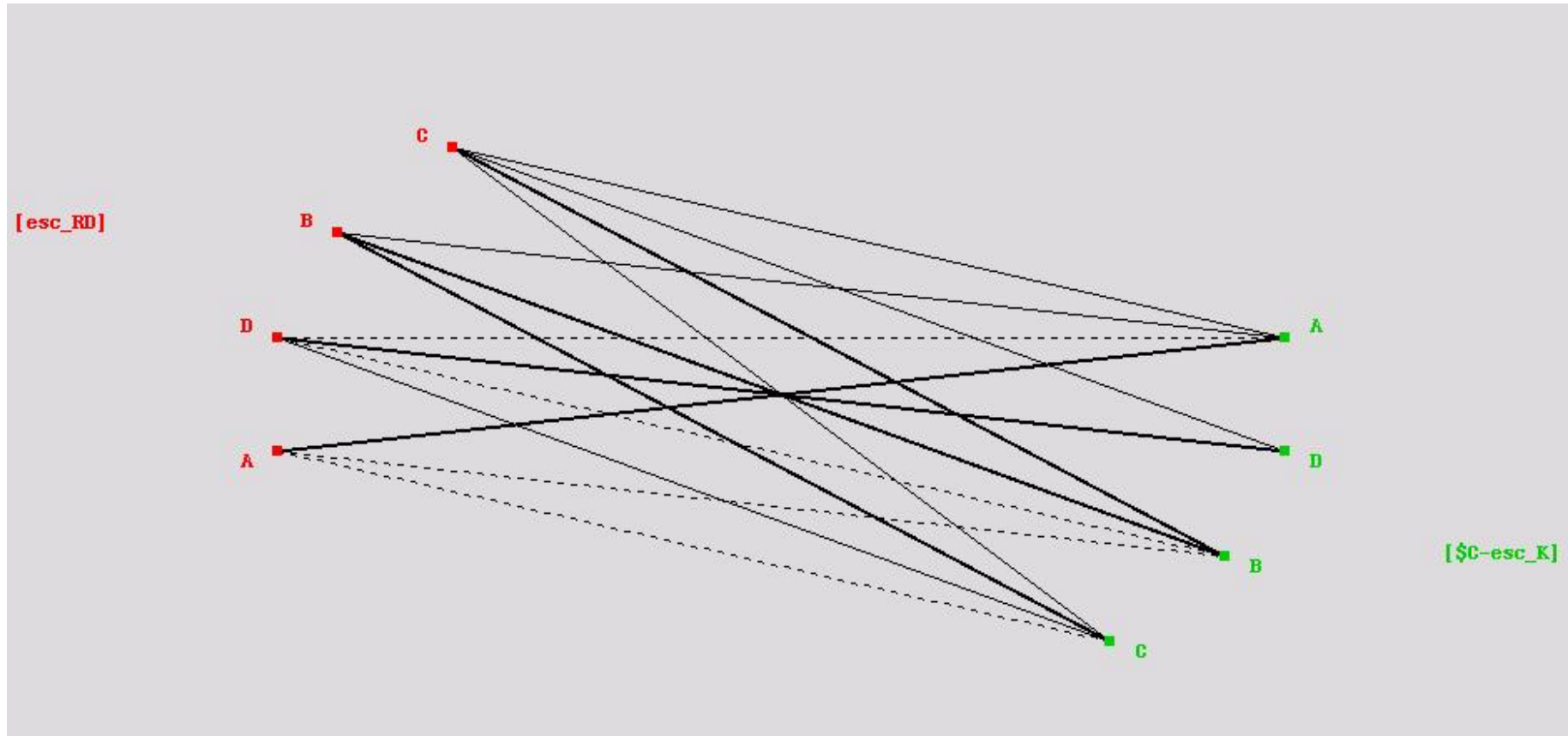
esc\_K == A <= esc\_UI == Muito Elevado & esc\_RD == A (238:18.45%, 0.74)  
esc\_K == A <= esc\_RD == A (274:21.24%, 0.69)  
esc\_K == D <= esc\_UI == Elevado & esc\_RD == D (242:18.76%, 0.76)  
esc\_K == A <= esc\_SL50 == D & esc\_RD == A (185:14.341%, 0.77)  
esc\_K == A <= esc\_SL50 == D & esc\_UI == Muito Elevado & esc\_RD == A (185:14.341%, 0.77)  
esc\_K == A <= esc\_SL50 == D (353:27.364%, 0.61)  
esc\_K == D <= esc\_RD == D (307:23.798%, 0.69)  
esc\_K == D <= esc\_SL50 == A & esc\_RD == D (124:9.612%, 0.85)  
esc\_K == B <= esc\_UI == Muito Elevado & esc\_RD == B (217:16.822%, 0.67)  
esc\_K == C <= esc\_SL50 == A & esc\_RD == B (71:5.504%, 0.96)  
esc\_K == D <= esc\_SL50 == A & esc\_UI == Elevado & esc\_RD == D (92:7.132%, 0.89)  
esc\_K == B <= esc\_SL50 == C & esc\_UI == Muito Elevado (269:20.853%, 0.61)  
esc\_K == D <= esc\_SL50 == B & esc\_UI == Elevado & esc\_RD == D (130:10.078%, 0.75)  
esc\_K == D <= esc\_SL50 == B & esc\_RD == D (132:10.233%, 0.75)  
esc\_K == C <= esc\_UI == Elevado & esc\_RD == B (173:13.411%, 0.66)  
esc\_K == C <= esc\_SL50 == A & esc\_UI == Elevado & esc\_RD == B (47:3.643%, 0.96)  
esc\_K == D <= esc\_SL50 == B & esc\_UI == Elevado (281:21.783%, 0.52)

esc\_K == D <= esc\_SL50 == B (288:22.326%, 0.51)  
esc\_K == C <= esc\_SL50 == B & esc\_UI == Elevado & esc\_RD == B (82:6.357%, 0.67)  
esc\_K == D <= esc\_SL50 == A & esc\_UI == Elevado (177:13.721%, 0.55)  
esc\_K == C <= esc\_SL50 == B & esc\_RD == B (85:6.589%, 0.65)  
esc\_K == C <= esc\_UI == Medio & esc\_RD == B (24:1.86%, 0.96)  
esc\_K == C <= esc\_SL50 == A & esc\_UI == Medio & esc\_RD == B (24:1.86%, 0.96)  
esc\_K == A <= esc\_SL50 == C & esc\_RD == A (74:5.736%, 0.61)  
esc\_K == C <= esc\_UI == Medio (78:6.047%, 0.62)  
esc\_K == C <= esc\_SL50 == A & esc\_UI == Medio (75:5.814%, 0.61)  
esc\_K == A <= esc\_SL50 == C & esc\_UI == Muito Elevado & esc\_RD == A (53:4.109%, 0.6)  
esc\_K == B <= esc\_SL50 == D & esc\_UI == Muito Elevado & esc\_RD == B (94:7.287%, 0.53)  
esc\_K == B <= esc\_SL50 == D & esc\_RD == B (95:7.364%, 0.53)  
esc\_K == C <= esc\_SL50 == A & esc\_UI == Elevado & esc\_RD == A (9:0.698%, 1.0)  
esc\_K == C <= esc\_SL50 == C & esc\_UI == Elevado & esc\_RD == D (20:1.55%, 0.75)  
esc\_K == C <= esc\_SL50 == C & esc\_RD == D (41:3.178%, 0.59)  
esc\_K == A <= esc\_SL50 == C & esc\_UI == Elevado & esc\_RD == A (21:1.628%, 0.62)  
esc\_K == C <= esc\_SL50 == B & esc\_RD == A (5:0.388%, 1.0)  
esc\_K == C <= esc\_SL50 == B & esc\_UI == Elevado & esc\_RD == A (5:0.388%, 1.0)  
esc\_K == B <= esc\_SL50 == B & esc\_UI == Muito Elevado (4:0.31%, 0.75)  
esc\_K == C <= esc\_SL50 == B & esc\_UI == Medio & esc\_RD == C (1:0.078%, 1.0)

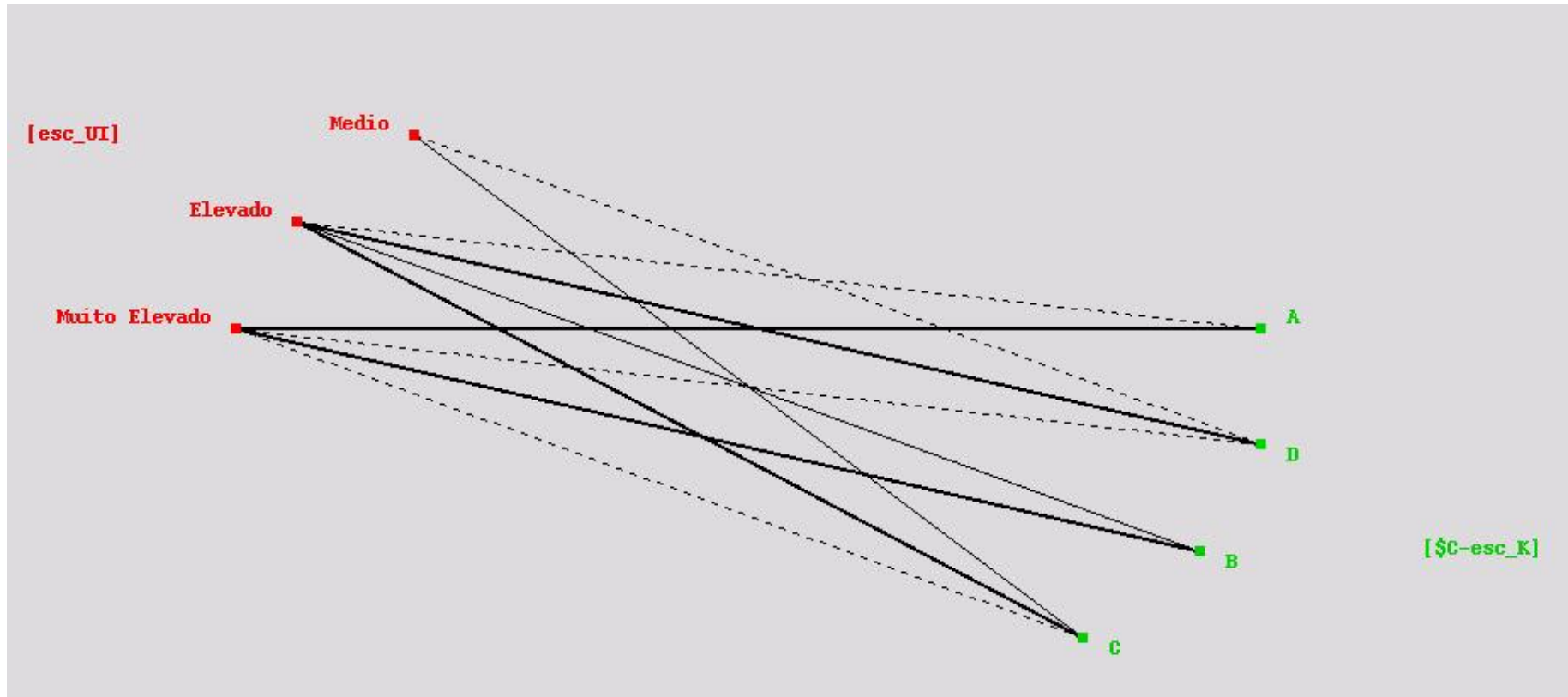
### Esc\_K vs. Esc\_SL50



### Esc\_K vs. Esc\_RD



### Esc\_K vs. Esc\_UI





## **ANEXO V**

**(Resultados obtidos relativamente  
à propriedade ME)**

**Regras geradas com o modelo**  
**C5.0 para o campo %ME**

Rules for A:

Rule #1 for A:

if esc\_RD == A  
and esc\_\_b == C  
then -> A

Rule #2 for A:

if esc\_RD == A  
and esc\_\_b == A  
and esc\_ST == C  
then -> A

Rule #3 for A:

if esc\_RD == A  
and esc\_\_b == A  
and esc\_ST == D  
then -> A

Rule #4 for A:

if esc\_RD == A  
and esc\_\_b == B  
and esc\_ST == A  
then -> A

Rule #5 for A:

if esc\_RD == A  
and esc\_\_b == B

and esc\_ST == D  
then -> A

Rule #6 for A:

if esc\_RD == A  
and esc\_\_b == D  
and esc\_ST == D  
then -> A

Rule #7 for A:

if esc\_RD == D  
and esc\_micronaire == Fina  
and esc\_PSI == D  
then -> A

Rule #8 for A:

if esc\_RD == D  
and esc\_micronaire ==  
Grossa  
and esc\_\_b == A  
then -> A

Rule #9 for A:

if esc\_RD == D  
and esc\_micronaire ==  
Grossa  
and esc\_\_b == [B D]  
then -> A

Rule #10 for A:

if esc\_RD == B  
and esc\_PSI == C  
then -> A

Rule #11 for A:

if esc\_RD == B  
and esc\_PSI == D  
then -> A

Rule #12 for A:

if esc\_RD == B  
and esc\_PSI == B  
and esc\_micronaire == Fina  
then -> A

Rule #13 for A:

if esc\_RD == B  
and esc\_PSI == B  
and esc\_micronaire == Muito Fina  
then -> A

Rule #14 for A:

if esc\_RD == B  
and esc\_PSI == A  
and esc\_\_b == C  
then -> A

Rule #15 for A:

if esc\_RD == B  
and esc\_PSI == A

```

and esc_b == D
  then -> A

Rule #16 for A:
  if esc_RD == B
  and esc_PSI == A
  and esc_b == B
and esc_micronaire == Fina
  then -> A

Rule #17 for A:
  if esc_RD == C
  and esc_ST == D
  then -> A

Rule #18 for A:
  if esc_RD == C
  and esc_ST == C
and esc_micronaire == Fina
  then -> A

Rule #19 for A:
  if esc_RD == C
  and esc_ST == C
and esc_micronaire == Muito
  Fina
  then -> A

Rule #20 for A:
  if esc_RD == C

```

```

and esc_ST == C
and esc_micronaire ==
  Grossa
  and esc_b == A
  then -> A

```

```

Rules for B:
Rule #1 for B:
  if esc_RD == A
  and esc_b == A
  and esc_ST == A
  then -> B

```

```

Rule #2 for B:
  if esc_RD == A
  and esc_b == A
  and esc_ST == B
  then -> B

```

```

Rule #3 for B:
  if esc_RD == A
  and esc_b == B
  and esc_ST == B
  then -> B

```

```

Rule #4 for B:
  if esc_RD == A
  and esc_b == B
  and esc_ST == C
  then -> B

```

```

Rule #5 for B:
  if esc_RD == D
  and esc_micronaire == Media
  then -> B

```

```

Rule #6 for B:
  if esc_RD == D
  and esc_micronaire == Fina
  and esc_PSI == C
  then -> B

```

```

Rule #7 for B:
  if esc_RD == D
  and esc_micronaire == Grossa
  and esc_b == C
  then -> B

```

```

Rule #8 for B:
  if esc_RD == B
  and esc_PSI == B
  and esc_micronaire == Grossa
  then -> B

```

```

Rule #9 for B:
  if esc_RD == B
  and esc_PSI == B
  and esc_micronaire == Media
  then -> B

```

<p>Rule #10 for B:  if esc_RD == B  and esc_PSI == A  and esc__b == A  then -&gt; B</p> <p>Rule #11 for B:  if esc_RD == B  and esc_PSI == A  and esc__b == B  and esc_micronaire ==  [Grossa Muito Fina]  then -&gt; B</p> <p>Rule #12 for B:  if esc_RD == B  and esc_PSI == A  and esc__b == B  and esc_micronaire ==  Media  then -&gt; B</p> <p>Rule #13 for B:  if esc_RD == C  and esc_ST == A  then -&gt; B</p> <p>Rule #14 for B:  if esc_RD == C  and esc_ST == B</p>	<p>then -&gt; B</p> <p>Rule #15 for B:  if esc_RD == C  and esc_ST == C  and esc_micronaire ==  Media  then -&gt; B</p> <p>Rule #16 for B:  if esc_RD == C  and esc_ST == C  and esc_micronaire ==  Grossa  and esc__b == B  then -&gt; B</p> <p>Rule #17 for B:  if esc_RD == C  and esc_ST == C  and esc_micronaire ==  Grossa  and esc__b == C  then -&gt; B</p> <p>Rule #18 for B:  if esc_RD == C  and esc_ST == C  and esc_micronaire ==  Grossa</p>	<p>and esc__b == D  then -&gt; B</p> <p>Rules for C:  Rule #1 for C:  if esc_RD == D  and esc_micronaire == Muito Fina  then -&gt; C</p> <p>Rule #2 for C:  if esc_RD == D  and esc_micronaire == Fina  and esc_PSI == A  then -&gt; C</p> <p>Rule #3 for C:  if esc_RD == D  and esc_micronaire == Fina  and esc_PSI == B  then -&gt; C</p> <p>Rules for D:  Rule #1 for D:  if esc_RD == A  and esc__b == D  and esc_ST == [A B]  then -&gt; D</p> <p>Rule #2 for D:  if esc_RD == A</p>
---	--	--

```
and esc__b == D
and esc_ST == C
  then -> D
```

Default : -> B

### **Regras geradas com o modelo GRI para o campo %ME**

esc\_ME == D <= esc\_ST == C & esc\_\_b == D & esc\_RD == A  
(136:10.543%, 0.92)  
esc\_ME == C <= esc\_ST == A & esc\_RD == D (164:12.713%,  
0.88)  
esc\_ME == C <= esc\_ST == A & esc\_PSI == A & esc\_RD == D  
(164:12.713%, 0.88)  
esc\_ME == C <= esc\_PSI == A & esc\_RD == D (177:13.721%,  
0.85)  
esc\_ME == D <= esc\_\_b == D & esc\_RD == A (202:15.659%,  
0.73)  
esc\_ME == D <= esc\_ST == C & esc\_RD == A (156:12.093%, 0.8)  
esc\_ME == A <= esc\_ST == D & esc\_PSI == D (276:21.395%,  
0.92)  
esc\_ME == D <= esc\_PSI == C & esc\_\_b == D & esc\_RD == A  
(108:8.372%, 0.93)  
esc\_ME == A <= esc\_ST == D (277:21.473%, 0.92)  
esc\_ME == C <= esc\_micronaire == Fina & esc\_RD == D  
(140:10.853%, 0.85)  
esc\_ME == C <= esc\_micronaire == Fina & esc\_PSI == A &  
esc\_RD == D (103:7.984%, 0.95)  
esc\_ME == C <= esc\_micronaire == Fina & esc\_ST == A &  
esc\_RD == D (95:7.364%, 0.97)  
esc\_ME == D <= esc\_ST == C & esc\_\_b == D (178:13.798%, 0.7)  
esc\_ME == D <= esc\_ST == C & esc\_PSI == C & esc\_RD == A  
(127:9.845%, 0.79)  
esc\_ME == C <= esc\_RD == D (307:23.798%, 0.56)  
esc\_ME == D <= esc\_PSI == C & esc\_RD == A (128:9.922%,  
0.78)  
esc\_ME == D <= esc\_micronaire == Fina & esc\_PSI == C &  
esc\_RD == A (76:5.891%, 0.96)

esc\_ME == C <= esc\_micronaire == Fina & esc\_PSI == A  
(119:9.225%, 0.82)  
esc\_ME == D <= esc\_RD == A (274:21.24%, 0.54)  
esc\_ME == D <= esc\_micronaire == Fina & esc\_ST == C & esc\_RD  
== A (77:5.969%, 0.95)  
esc\_ME == C <= esc\_micronaire == Fina & esc\_ST == A & esc\_PSI  
== A (108:8.372%, 0.85)  
esc\_ME == D <= esc\_micronaire == Fina & esc\_\_b == D & esc\_RD  
== A (91:7.054%, 0.88)  
esc\_ME == C <= esc\_micronaire == Fina & esc\_ST == A  
(109:8.45%, 0.84)  
esc\_ME == D <= esc\_micronaire == Fina & esc\_PSI == C & esc\_\_b  
== D (82:6.357%, 0.89)  
esc\_ME == D <= esc\_ST == C & esc\_PSI == C & esc\_\_b == D  
(148:11.473%, 0.68)  
esc\_ME == D <= esc\_micronaire == Fina & esc\_ST == C & esc\_\_b  
== D (83:6.434%, 0.88)  
esc\_ME == D <= esc\_micronaire == Fina & esc\_RD == A  
(98:7.597%, 0.82)  
esc\_ME == D <= esc\_PSI == C & esc\_\_b == D (151:11.705%, 0.66)  
esc\_ME == A <= esc\_micronaire == Media & esc\_ST == D &  
esc\_PSI == D (148:11.473%, 0.95)  
esc\_ME == C <= esc\_PSI == A & esc\_\_b == D & esc\_RD == D  
(57:4.419%, 1.0)  
esc\_ME == C <= esc\_ST == A & esc\_\_b == D & esc\_RD == D  
(57:4.419%, 1.0)  
esc\_ME == A <= esc\_micronaire == Media & esc\_ST == D  
(149:11.55%, 0.95)  
esc\_ME == C <= esc\_ST == A & esc\_\_b == D (61:4.729%, 0.93)  
esc\_ME == C <= esc\_ST == A & esc\_PSI == A & esc\_\_b == D  
(61:4.729%, 0.93)  
esc\_ME == C <= esc\_PSI == A & esc\_\_b == D (62:4.806%, 0.92)

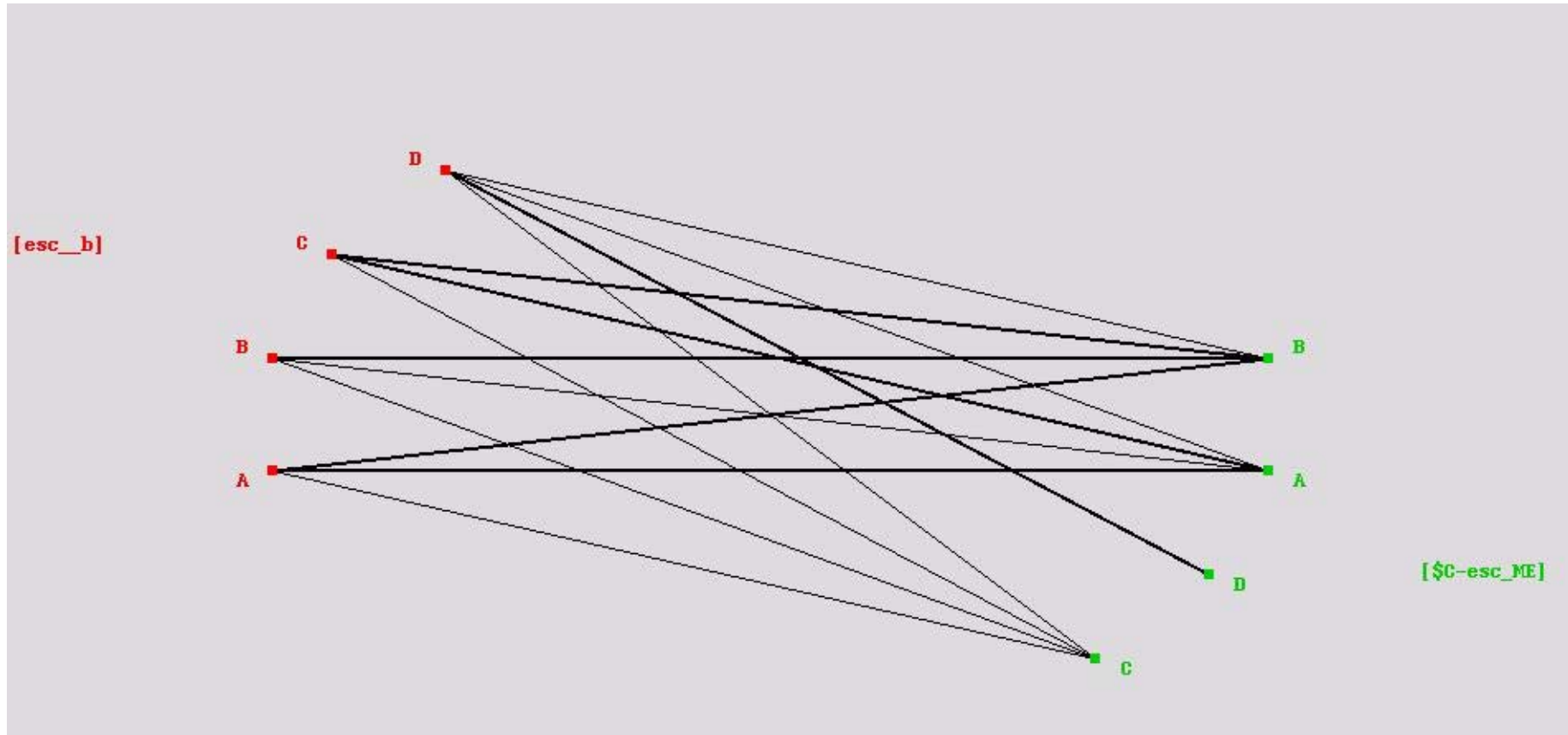
esc\_ME == D <= esc\_micronaire == Fina & esc\_\_b == D  
 (127:9.845%, 0.63)  
 esc\_ME == B <= esc\_micronaire == Media & esc\_ST == B  
 (234:18.14%, 0.83)  
 esc\_ME == C <= esc\_micronaire == Muito Fina & esc\_PSI == A  
 (45:3.488%, 1.0)  
 esc\_ME == C <= esc\_micronaire == Muito Fina & esc\_ST == A  
 (45:3.488%, 1.0)  
 esc\_ME == D <= esc\_micronaire == Fina & esc\_ST == C &  
 esc\_PSI == C (114:8.837%, 0.64)  
 esc\_ME == C <= esc\_\_b == D & esc\_RD == D (68:5.271%, 0.84)  
 esc\_ME == D <= esc\_micronaire == Fina & esc\_PSI == C  
 (118:9.147%, 0.62)  
 esc\_ME == D <= esc\_micronaire == Fina & esc\_ST == C  
 (127:9.845%, 0.57)  
 esc\_ME == C <= esc\_ST == A & esc\_\_b == C & esc\_RD == D  
 (41:3.178%, 0.95)  
 esc\_ME == C <= esc\_PSI == A & esc\_\_b == C & esc\_RD == D  
 (42:3.256%, 0.93)  
 esc\_ME == B <= esc\_ST == B (319:24.729%, 0.71)  
 esc\_ME == D <= esc\_micronaire == Muito Fina & esc\_ST == C &  
 esc\_RD == A (31:2.403%, 1.0)  
 esc\_ME == D <= esc\_micronaire == Muito Fina & esc\_ST == C &  
 esc\_\_b == D (31:2.403%, 1.0)  
 esc\_ME == D <= esc\_micronaire == Muito Fina & esc\_\_b == D &  
 esc\_RD == A (46:3.566%, 0.85)  
 esc\_ME == D <= esc\_micronaire == Muito Fina & esc\_RD == A  
 (48:3.721%, 0.81)  
 esc\_ME == B <= esc\_micronaire == Media (741:57.442%, 0.59)  
 esc\_ME == D <= esc\_micronaire == Muito Fina & esc\_ST == C  
 (40:3.101%, 0.78)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_PSI == D  
 (75:5.814%, 0.89)

esc\_ME == D <= esc\_PSI == B & esc\_\_b == D & esc\_RD == A  
 (28:2.171%, 0.89)  
 esc\_ME == D <= esc\_ST == C & esc\_PSI == B & esc\_\_b == D  
 (28:2.171%, 0.89)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_ST == D & esc\_PSI  
 == D (64:4.961%, 0.89)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_ST == D  
 (64:4.961%, 0.89)  
 esc\_ME == A <= esc\_micronaire == Grossa & esc\_PSI == D  
 (37:2.868%, 1.0)  
 esc\_ME == C <= esc\_\_b == B & esc\_RD == D (73:5.659%, 0.56)  
 esc\_ME == C <= esc\_ST == A & esc\_\_b == A & esc\_RD == D  
 (39:3.023%, 0.72)  
 esc\_ME == D <= esc\_PSI == B & esc\_\_b == D (36:2.791%, 0.69)  
 esc\_ME == A <= esc\_micronaire == Grossa & esc\_ST == D  
 (33:2.558%, 1.0)  
 esc\_ME == C <= esc\_PSI == A & esc\_\_b == B & esc\_RD == D  
 (36:2.791%, 0.72)  
 esc\_ME == C <= esc\_PSI == A & esc\_\_b == A & esc\_RD == D  
 (42:3.256%, 0.67)  
 esc\_ME == D <= esc\_ST == C & esc\_PSI == B (38:2.946%, 0.66)  
 esc\_ME == C <= esc\_\_b == C & esc\_RD == D (78:6.047%, 0.5)  
 esc\_ME == C <= esc\_ST == A & esc\_\_b == B & esc\_RD == D  
 (27:2.093%, 0.74)  
 esc\_ME == A <= esc\_micronaire == Muito Fina & esc\_RD == B  
 (20:1.55%, 1.0)  
 esc\_ME == A <= esc\_ST == C & esc\_RD == B (71:5.504%, 0.73)  
 esc\_ME == D <= esc\_micronaire == Media & esc\_PSI == B &  
 esc\_\_b == D (20:1.55%, 0.7)  
 esc\_ME == C <= esc\_ST == B & esc\_\_b == B & esc\_RD == D  
 (38:2.946%, 0.55)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_ST == C & esc\_RD  
 == B (24:1.86%, 0.92)

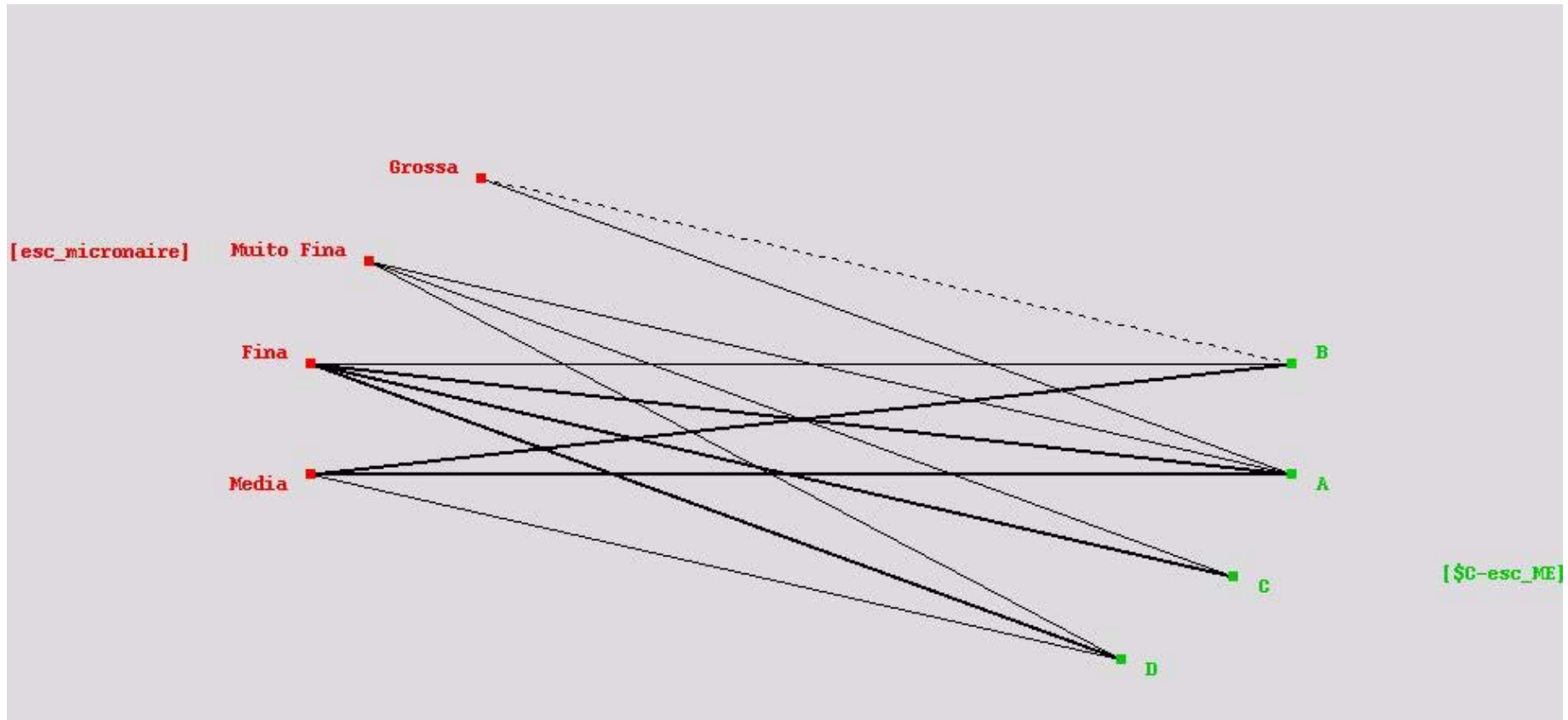
esc\_ME == A <= esc\_micronaire == Grossa (52:4.031%, 0.75)  
 esc\_ME == D <= esc\_micronaire == Media & esc\_ST == C &  
 esc\_PSI == B (25:1.938%, 0.56)  
 esc\_ME == C <= esc\_PSI == B & esc\_\_b == B & esc\_RD == D  
 (30:2.326%, 0.5)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_ST == C &  
 esc\_PSI == D (11:0.853%, 0.91)  
 esc\_ME == B <= esc\_micronaire == Grossa & esc\_\_b == D  
 (7:0.543%, 1.0)  
 esc\_ME == B <= esc\_micronaire == Grossa & esc\_PSI == C  
 (10:0.775%, 0.9)  
 esc\_ME == B <= esc\_micronaire == Grossa & esc\_ST == C &  
 esc\_PSI == C (10:0.775%, 0.9)  
 esc\_ME == B <= esc\_micronaire == Grossa & esc\_PSI == B  
 (4:0.31%, 1.0)  
 esc\_ME == B <= esc\_micronaire == Grossa & esc\_ST == B  
 (4:0.31%, 1.0)  
 esc\_ME == B <= esc\_micronaire == Fina & esc\_ST == B &  
 esc\_PSI == C (4:0.31%, 1.0)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_ST == B & esc\_\_b  
 == C (7:0.543%, 0.86)  
 esc\_ME == B <= esc\_micronaire == Fina & esc\_ST == A &  
 esc\_RD == C (7:0.543%, 0.86)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_ST == C &  
 esc\_PSI == B (2:0.155%, 1.0)  
 esc\_ME == A <= esc\_micronaire == Fina & esc\_ST == A &  
 esc\_RD == B (7:0.543%, 0.71)  
 esc\_ME == B <= esc\_micronaire == Fina & esc\_PSI == C &  
 esc\_RD == D (5:0.388%, 0.8)  
 esc\_ME == B <= esc\_micronaire == Fina & esc\_PSI == A &  
 esc\_RD == C (9:0.698%, 0.67)  
 esc\_ME == B <= esc\_micronaire == Fina & esc\_PSI == C &  
 esc\_\_b == A (15:1.163%, 0.6)

esc\_ME == B <= esc\_micronaire == Fina & esc\_\_b == B & esc\_RD  
 == A (4:0.31%, 0.75)  
 esc\_ME == B <= esc\_micronaire == Fina & esc\_PSI == B & esc\_RD  
 == A (4:0.31%, 0.75)

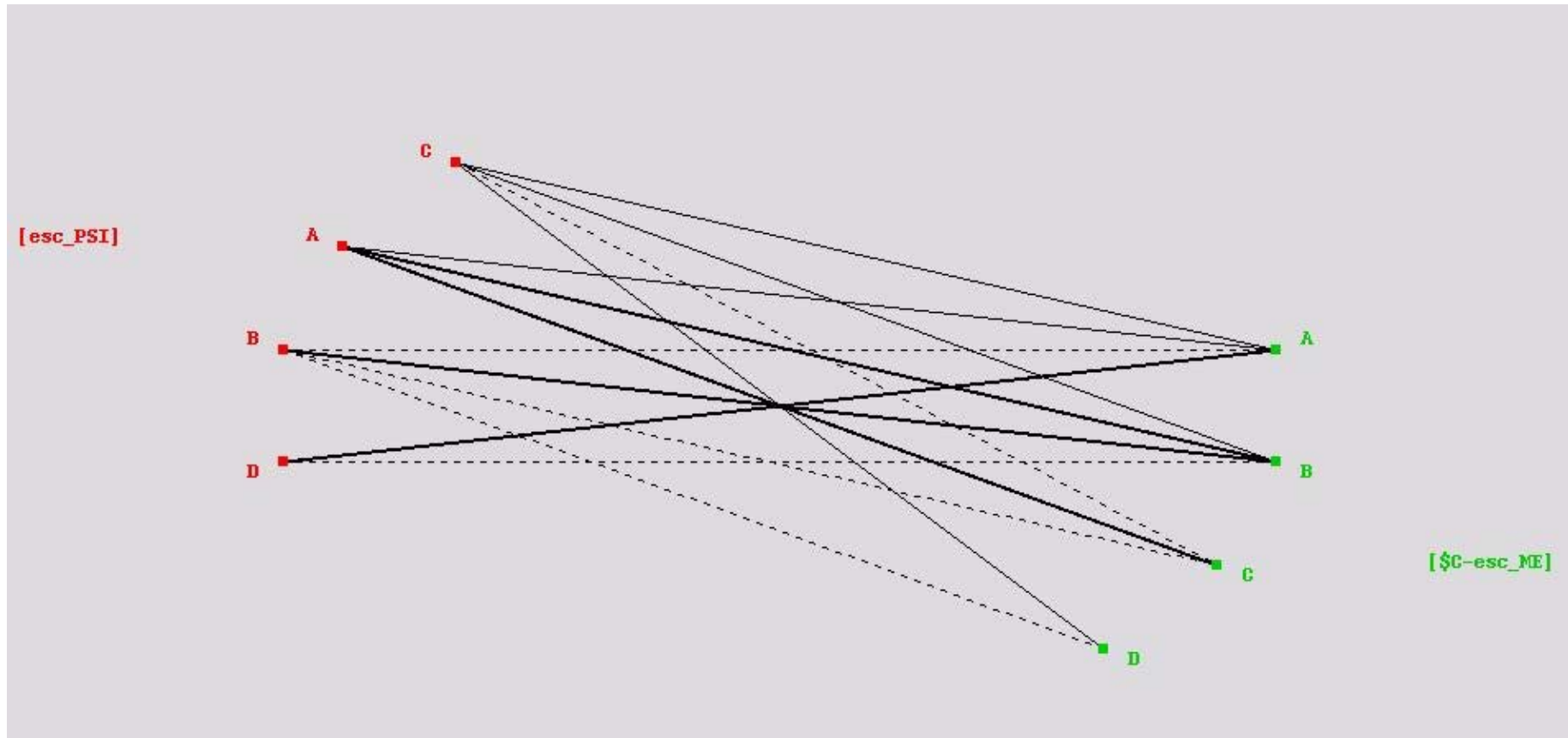
Esc\_ME vs. Esc\_+b



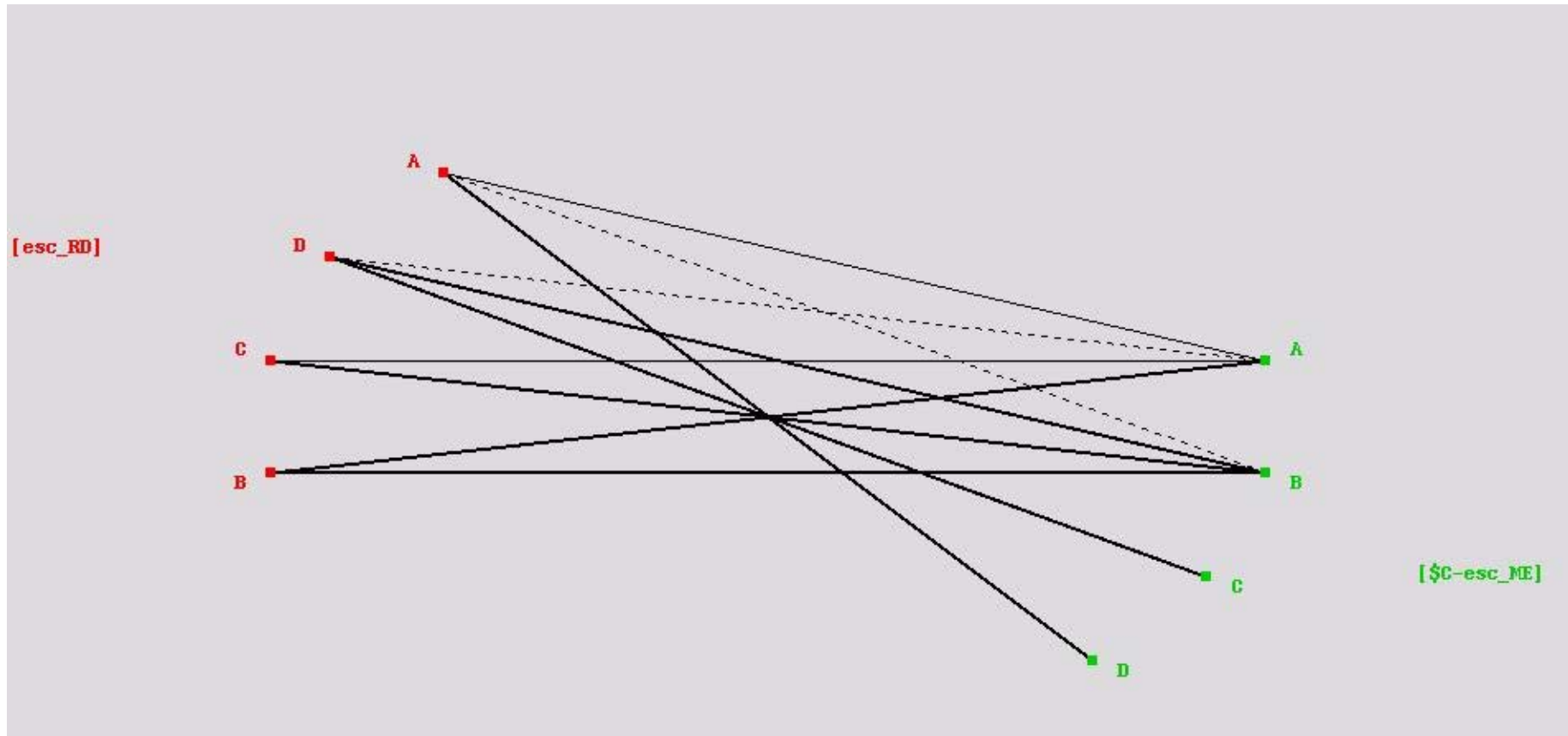
### Esc\_ME vs. Esc\_Micronaire



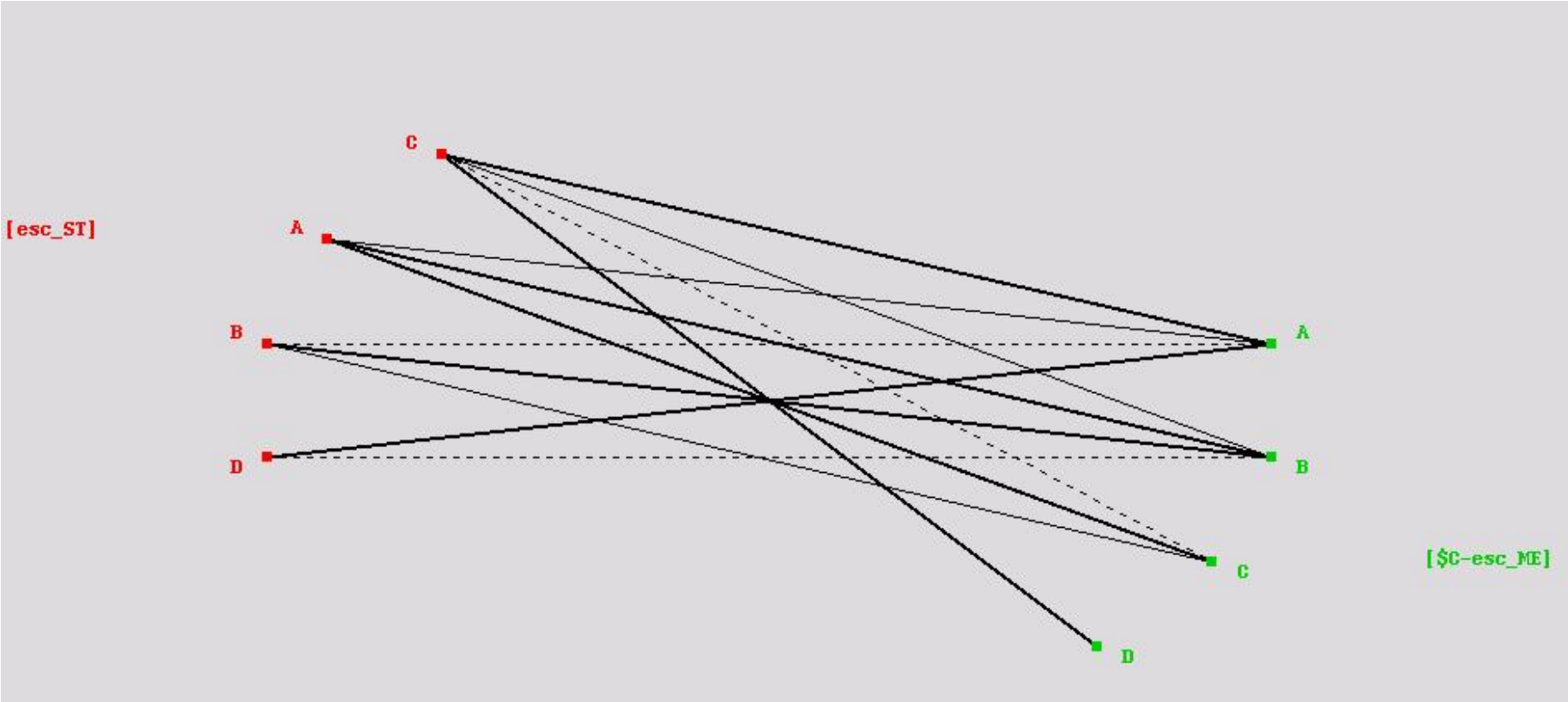
Esc\_ME vs. Esc\_PSI



### Esc\_ME vs. Esc\_RD



Esc\_ME vs. Esc\_ST





## **ANEXO VI**

**(Resultados obtidos relativamente  
à propriedade Mg)**

**Regras geradas com o modelo  
C5.0 para o campo Mg**

Rules for A:

Rule #1 for A:

if esc\_SL2\_5 == A  
and esc\_PSI == A  
then -> A

Rule #2 for A:

if esc\_SL2\_5 == A  
and esc\_PSI == B  
then -> A

Rule #3 for A:

if esc\_SL2\_5 == B  
and esc\_PSI == A  
then -> A

Rule #4 for A:

if esc\_SL2\_5 == B  
and esc\_PSI == B  
then -> A

Rules for B:

Rule #1 for B:

if esc\_SL2\_5 == A  
and esc\_PSI == C  
then -> B

Rule #2 for B:

if esc\_SL2\_5 == A  
and esc\_PSI == D  
then -> B

Rule #3 for B:

if esc\_SL2\_5 == C  
and esc\_PSI == A  
then -> B

Rule #4 for B:

if esc\_SL2\_5 == C  
and esc\_PSI == B  
then -> B

Rules for C:

Rule #1 for C:

if esc\_SL2\_5 == B  
and esc\_PSI == C  
then -> C

Rule #2 for C:

if esc\_SL2\_5 == C  
and esc\_PSI == C  
then -> C

Rules for D:

Rule #1 for D:

if esc\_SL2\_5 == D  
then -> D

Rule #2 for D:

if esc\_SL2\_5 == B  
and esc\_PSI == D  
then -> D

Rule #3 for D:

if esc\_SL2\_5 == C  
and esc\_PSI == D  
then -> D

Default : -> A

### **Regras geradas com o modelo GRI para o campo Mg**

esc\_Mg == A <= esc\_PSI == A (351:27.209%, 0.83)

esc\_Mg == D <= esc\_SL2\_5 == D (309:23.953%, 0.66)

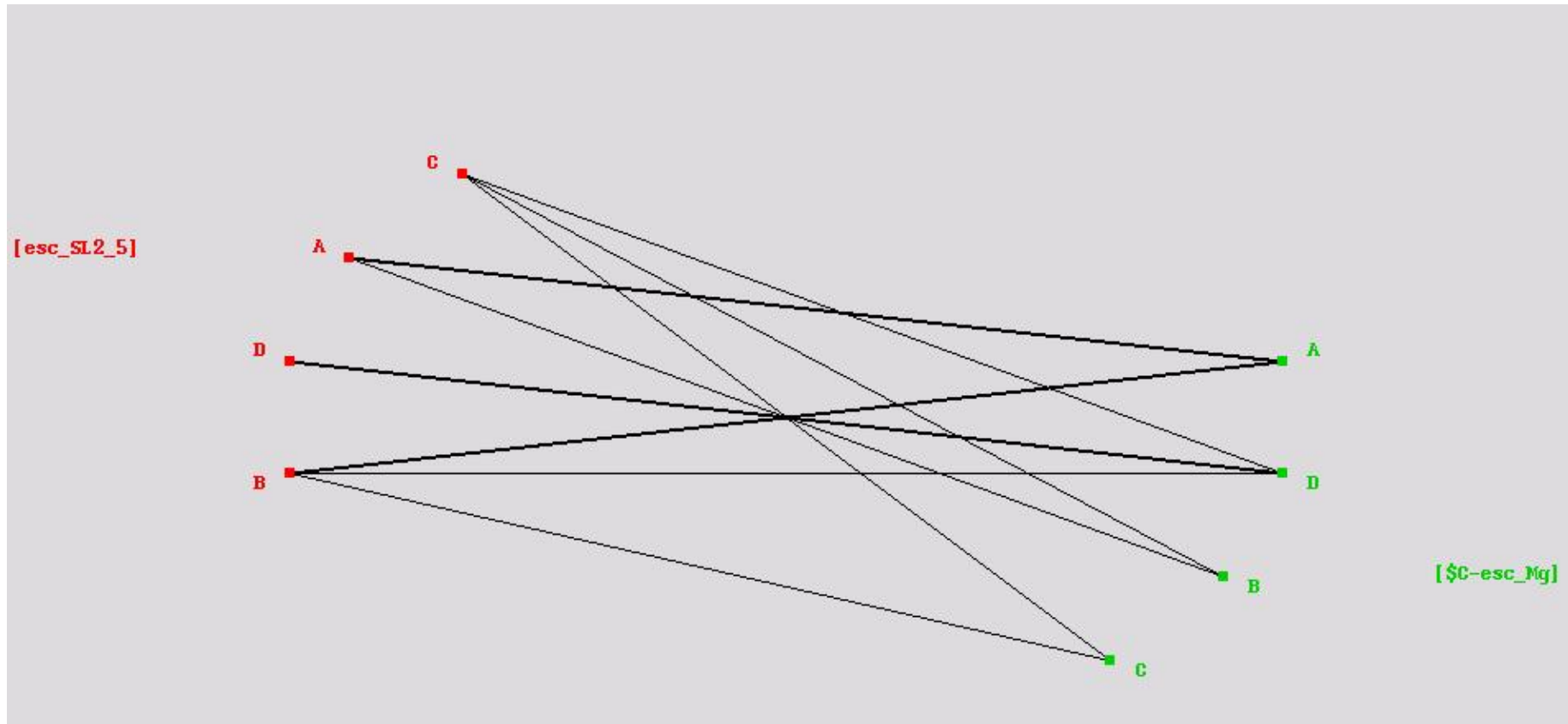
esc\_Mg == B <= esc\_PSI == C & esc\_SL2\_5 == A (70:5.426%, 0.87)

esc\_Mg == D <= esc\_PSI == C & esc\_SL2\_5 == D (115:8.915%, 0.77)

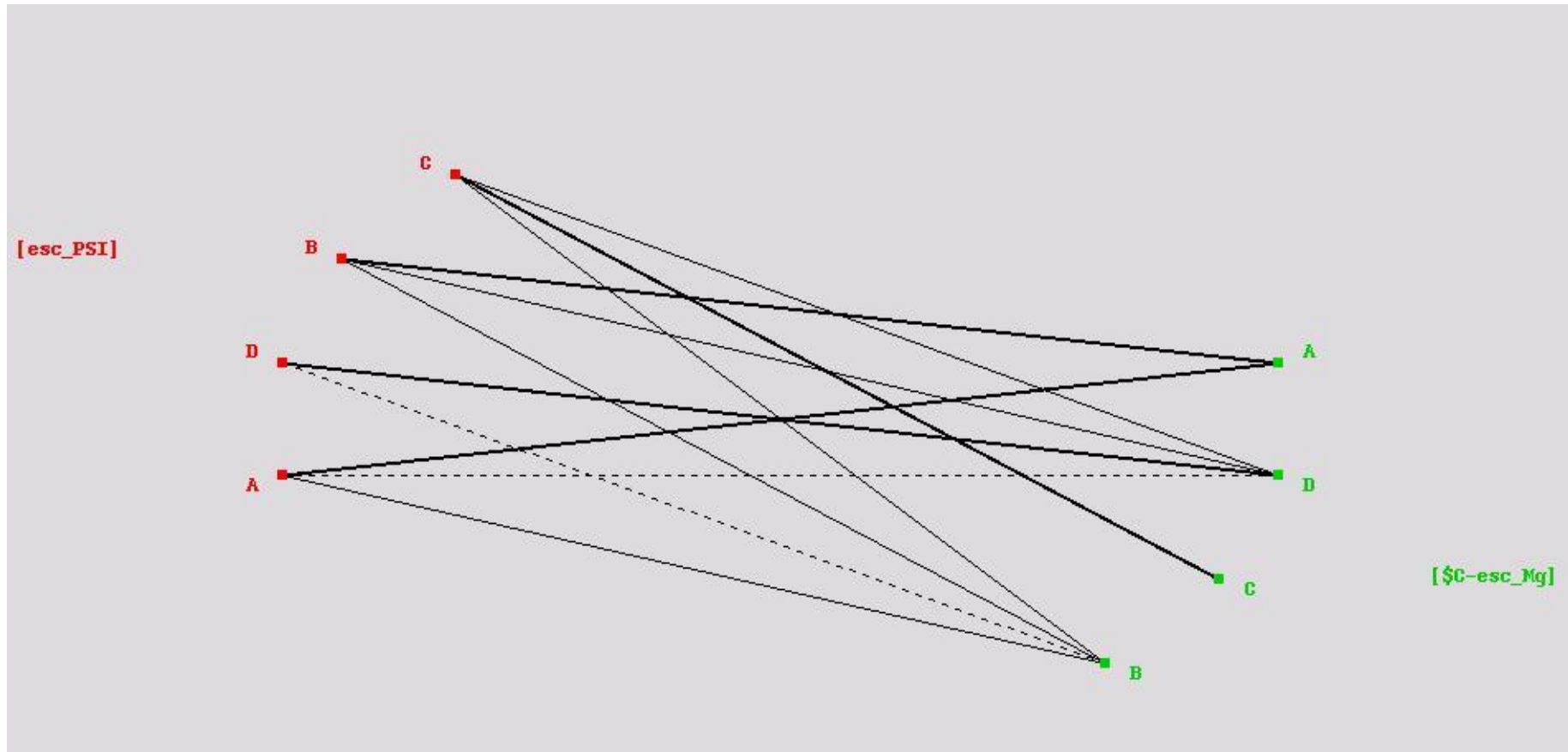
esc\_Mg == D <= esc\_PSI == D (309:23.953%, 0.51)

esc\_Mg == D <= esc\_PSI == B & esc\_SL2\_5 == D (29:2.248%, 0.86)

### Esc\_Mg vs. Esc\_SL2.5



### Esc\_Mg vs. Esc\_PSI





## **ANEXO VII**

**(Resultados obtidos relativamente  
à propriedade P)**

**Regras geradas com o modelo  
C5.0 para o campo P**

Rules for A:

Rule #1 for A:

if esc\_SL50 == A  
and esc\_RD == D  
and esc\_PSI == A  
then -> A

Rule #2 for A:

if esc\_SL50 == A  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_b == B  
then -> A

Rule #3 for A:

if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == A  
then -> A

Rule #4 for A:

if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == D  
then -> A

Rule #5 for A:

if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_b == A  
and esc\_EL == Baixo  
then -> A

Rule #6 for A:

if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_b == A  
and esc\_EL == [Elevado  
Muito\_Baixo Muito\_Elevado]  
then -> A

Rule #7 for A:

if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_b == B  
and esc\_EL == [Elevado  
Muito\_Baixo Muito\_Elevado]  
then -> A

Rule #8 for A:

if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_b == B

and esc\_EL == Medio  
then -> A

Rules for B:

Rule #1 for B:

if esc\_SL50 == A  
and esc\_RD == A  
then -> B

Rule #2 for B:

if esc\_SL50 == A  
and esc\_RD == B  
then -> B

Rule #3 for B:

if esc\_SL50 == A  
and esc\_RD == C  
then -> B

Rule #4 for B:

if esc\_SL50 == A  
and esc\_RD == D  
and esc\_PSI == C  
then -> B

Rule #5 for B:

if esc\_SL50 == A  
and esc\_RD == D  
and esc\_PSI == D  
then -> B

Rule #6 for B:  
if esc\_SL50 == A  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_\_b == A  
then -> B

Rule #7 for B:  
if esc\_SL50 == A  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_\_b == C  
then -> B

Rule #8 for B:  
if esc\_SL50 == A  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_\_b == D  
then -> B

Rule #9 for B:  
if esc\_SL50 == D  
and esc\_PSI == A  
then -> B

Rule #10 for B:  
if esc\_SL50 == D  
and esc\_PSI == B

and esc\_RD == C  
then -> B

Rule #11 for B:  
if esc\_SL50 == B  
and esc\_RD == A  
then -> B

Rule #12 for B:  
if esc\_SL50 == B  
and esc\_RD == C  
then -> B

Rule #13 for B:  
if esc\_SL50 == B  
and esc\_RD == B  
and esc\_\_b == A  
then -> B

Rule #14 for B:  
if esc\_SL50 == B  
and esc\_RD == B  
and esc\_\_b == B  
then -> B

Rule #15 for B:  
if esc\_SL50 == B  
and esc\_RD == B  
and esc\_\_b == D  
then -> B

Rule #16 for B:  
if esc\_SL50 == B  
and esc\_RD == B  
and esc\_\_b == C  
and esc\_PSI == B  
then -> B

Rule #17 for B:  
if esc\_SL50 == B  
and esc\_RD == B  
and esc\_\_b == C  
and esc\_PSI == C  
then -> B

Rule #18 for B:  
if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == C  
then -> B

Rule #19 for B:  
if esc\_SL50 == B  
and esc\_RD == D  
and esc\_PSI == B  
and esc\_\_b == C  
then -> B

Rule #20 for B:  
if esc\_SL50 == B

<p>and esc_RD == D and esc_PSI == B and esc__b == D then -&gt; B</p> <p>Rule #21 for B: if esc_SL50 == B and esc_RD == D and esc_PSI == B and esc__b == A and esc_EL == Medio then -&gt; B</p> <p>Rule #22 for B: if esc_SL50 == B and esc_RD == D and esc_PSI == B and esc__b == B and esc_EL == Baixo then -&gt; B</p> <p>Rule #23 for B: if esc_SL50 == C and esc_PSI == A and esc_EL == Baixo then -&gt; B</p> <p>Rule #24 for B: if esc_SL50 == C and esc_PSI == A</p>	<p>and esc_EL == Muito_Baixo then -&gt; B</p> <p>Rule #25 for B: if esc_SL50 == C and esc_PSI == A and esc_EL == Muito_Elevado then -&gt; B</p> <p>Rule #26 for B: if esc_SL50 == C and esc_PSI == B and esc_EL == Baixo then -&gt; B</p> <p>Rule #27 for B: if esc_SL50 == C and esc_PSI == B and esc_EL == Muito_Baixo then -&gt; B</p> <p>Rule #28 for B: if esc_SL50 == C and esc_PSI == B and esc_EL == Medio and esc_RD == A then -&gt; B</p> <p>Rule #29 for B:</p>	<p>if esc_SL50 == C and esc_PSI == B and esc_EL == Medio and esc_RD == C then -&gt; B</p> <p>Rule #30 for B: if esc_SL50 == C and esc_PSI == B and esc_EL == Medio and esc_RD == D then -&gt; B</p> <p>Rule #31 for B: if esc_SL50 == C and esc_PSI == B and esc_EL == Elevado and esc__b == A and esc_RD == B then -&gt; B</p> <p>Rule #32 for B: if esc_SL50 == C and esc_PSI == B and esc_EL == Elevado and esc__b == A and esc_RD == D then -&gt; B</p> <p>Rule #33 for B:</p>
---	--	--

if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == D  
then -> B

Rule #34 for B:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == C  
and esc\_EL == Baixo  
then -> B

Rule #35 for B:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == C  
and esc\_EL == Medio  
then -> B

Rule #36 for B:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == C  
and esc\_EL == Muito\_Baixo  
then -> B

Rule #37 for B:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == C

and esc\_EL ==  
Muito\_Elevado  
then -> B

Rules for C:  
Rule #1 for C:  
if esc\_SL50 == D  
and esc\_PSI == B  
and esc\_RD == B  
then -> C

Rule #2 for C:  
if esc\_SL50 == D  
and esc\_PSI == C  
and esc\_b == B  
then -> C

Rule #3 for C:  
if esc\_SL50 == D  
and esc\_PSI == C  
and esc\_b == C  
then -> C

Rule #4 for C:  
if esc\_SL50 == D  
and esc\_PSI == C  
and esc\_b == D  
and esc\_RD == B  
then -> C

Rule #5 for C:  
if esc\_SL50 == D  
and esc\_PSI == C  
and esc\_b == D  
and esc\_RD == C  
then -> C

Rule #6 for C:  
if esc\_SL50 == D  
and esc\_PSI == D  
and esc\_EL == Elevado  
then -> C

Rule #7 for C:  
if esc\_SL50 == D  
and esc\_PSI == D  
and esc\_EL == Medio  
and esc\_b == B  
then -> C

Rule #8 for C:  
if esc\_SL50 == D  
and esc\_PSI == D  
and esc\_EL == Medio  
and esc\_b == C  
then -> C

Rule #9 for C:  
if esc\_SL50 == D  
and esc\_PSI == D

and esc\_EL == Medio  
and esc\_b == D  
then -> C

Rule #10 for C:  
if esc\_SL50 == B  
and esc\_RD == B  
and esc\_b == C  
and esc\_PSI == A  
then -> C

Rule #11 for C:  
if esc\_SL50 == B  
and esc\_RD == B  
and esc\_b == C  
and esc\_PSI == D  
then -> C

Rule #12 for C:  
if esc\_SL50 == C  
and esc\_PSI == A  
and esc\_EL == Elevado  
then -> C

Rule #13 for C:  
if esc\_SL50 == C  
and esc\_PSI == A  
and esc\_EL == Medio  
then -> C

Rule #14 for C:  
if esc\_SL50 == C  
and esc\_PSI == D  
and esc\_EL == Elevado  
then -> C

Rule #15 for C:  
if esc\_SL50 == C  
and esc\_PSI == D  
and esc\_EL == Medio  
then -> C

Rule #16 for C:  
if esc\_SL50 == C  
and esc\_PSI == D  
and esc\_EL ==  
Muito\_Elevado  
then -> C

Rule #17 for C:  
if esc\_SL50 == C  
and esc\_PSI == D  
and esc\_EL == Baixo  
and esc\_b == B  
then -> C

Rule #18 for C:  
if esc\_SL50 == C  
and esc\_PSI == D  
and esc\_EL == Baixo

and esc\_b == C  
then -> C

Rule #19 for C:  
if esc\_SL50 == C  
and esc\_PSI == B  
and esc\_EL == Muito\_Elevado  
then -> C

Rule #20 for C:  
if esc\_SL50 == C  
and esc\_PSI == B  
and esc\_EL == Medio  
and esc\_RD == B  
then -> C

Rule #21 for C:  
if esc\_SL50 == C  
and esc\_PSI == B  
and esc\_EL == Elevado  
and esc\_b == B  
then -> C

Rule #22 for C:  
if esc\_SL50 == C  
and esc\_PSI == B  
and esc\_EL == Elevado  
and esc\_b == C  
then -> C

Rule #23 for C:  
if esc\_SL50 == C  
and esc\_PSI == B  
and esc\_EL == Elevado  
and esc\_\_b == D  
then -> C

Rule #24 for C:  
if esc\_SL50 == C  
and esc\_PSI == B  
and esc\_EL == Elevado  
and esc\_\_b == A  
and esc\_RD == A  
then -> C

Rule #25 for C:  
if esc\_SL50 == C  
and esc\_PSI == B  
and esc\_EL == Elevado  
and esc\_\_b == A  
and esc\_RD == C  
then -> C

Rule #26 for C:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == B  
then -> C

Rule #27 for C:

if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == C  
and esc\_EL == Elevado  
then -> C

Rule #28 for C:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == A  
and esc\_EL == Baixo  
then -> C

Rule #29 for C:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == A  
and esc\_EL == Elevado  
then -> C

Rule #30 for C:  
if esc\_SL50 == C  
and esc\_PSI == C  
and esc\_RD == A  
and esc\_EL == Medio  
and esc\_\_b == C  
then -> C

Rules for D:  
Rule #1 for D:

if esc\_SL50 == D  
and esc\_PSI == B  
and esc\_RD == A  
then -> D

Rule #2 for D:  
if esc\_SL50 == D  
and esc\_PSI == B  
and esc\_RD == D  
then -> D

Rule #3 for D:  
if esc\_SL50 == D  
and esc\_PSI == C  
and esc\_\_b == A  
then -> D

Rule #4 for D:  
if esc\_SL50 == D  
and esc\_PSI == C  
and esc\_\_b == D  
and esc\_RD == A  
then -> D

Rule #5 for D:  
if esc\_SL50 == D  
and esc\_PSI == C  
and esc\_\_b == D  
and esc\_RD == D  
then -> D

<p>Rule #6 for D:  if esc_SL50 == D  and esc_PSI == D  and esc_EL == Baixo  then -&gt; D</p> <p>Rule #7 for D:  if esc_SL50 == D  and esc_PSI == D  and esc_EL == Muito_Baixo  then -&gt; D</p> <p>Rule #8 for D:  if esc_SL50 == D  and esc_PSI == D  and esc_EL ==  Muito_Elevado  then -&gt; D</p> <p>Rule #9 for D:  if esc_SL50 == D  and esc_PSI == D  and esc_EL == Medio  and esc_b == A  then -&gt; D</p> <p>Rule #10 for D:  if esc_SL50 == C  and esc_PSI == D</p>	<p>and esc_EL == Muito_Baixo  then -&gt; D</p> <p>Rule #11 for D:  if esc_SL50 == C  and esc_PSI == D  and esc_EL == Baixo  and esc_b == A  then -&gt; D</p> <p>Rule #12 for D:  if esc_SL50 == C  and esc_PSI == D  and esc_EL == Baixo  and esc_b == D  then -&gt; D</p> <p>Rule #13 for D:  if esc_SL50 == C  and esc_PSI == C  and esc_RD == A  and esc_EL == Muito_Baixo  then -&gt; D</p> <p>Rule #14 for D:  if esc_SL50 == C  and esc_PSI == C  and esc_RD == A  and esc_EL ==  Muito_Elevado</p>	<p>then -&gt; D</p> <p>Rule #15 for D:  if esc_SL50 == C  and esc_PSI == C  and esc_RD == A  and esc_EL == Medio  and esc_b == A  then -&gt; D</p> <p>Rule #16 for D:  if esc_SL50 == C  and esc_PSI == C  and esc_RD == A  and esc_EL == Medio  and esc_b == B  then -&gt; D</p> <p>Rule #17 for D:  if esc_SL50 == C  and esc_PSI == C  and esc_RD == A  and esc_EL == Medio  and esc_b == D  then -&gt; D</p> <p>Default : -&gt; B</p>
---	---	---

### **Regras geradas com o modelo GRI para o campo P**

esc\_P == A <= esc\_PSI == A & esc\_RD == D (177:13.721%, 0.85)  
esc\_P == D <= esc\_\_b == D & esc\_RD == A (202:15.659%, 0.86)  
esc\_P == A <= esc\_RD == D (307:23.798%, 0.56)  
esc\_P == D <= esc\_RD == A (274:21.24%, 0.69)  
esc\_P == D <= esc\_\_b == D & esc\_PSI == C & esc\_RD == A (108:8.372%, 0.93)  
esc\_P == A <= esc\_\_b == D & esc\_PSI == A & esc\_RD == D (57:4.419%, 1.0)  
esc\_P == A <= esc\_\_b == D & esc\_PSI == A (62:4.806%, 0.92)  
esc\_P == D <= esc\_PSI == C & esc\_RD == A (128:9.922%, 0.79)  
esc\_P == D <= esc\_\_b == D (322:24.961%, 0.54)  
esc\_P == A <= esc\_\_b == C & esc\_PSI == A & esc\_RD == D (42:3.256%, 0.93)  
esc\_P == C <= esc\_RD == B & esc\_SL50 == C (163:12.636%, 0.67)  
esc\_P == D <= esc\_\_b == A & esc\_PSI == D & esc\_SL50 == D (68:5.271%, 0.85)  
esc\_P == B <= esc\_SL50 == A & esc\_EL == Elevado (76:5.891%, 0.96)  
esc\_P == D <= esc\_\_b == A & esc\_PSI == D (105:8.14%, 0.71)  
esc\_P == B <= esc\_RD == B & esc\_SL50 == A (71:5.504%, 0.97)  
esc\_P == B <= esc\_\_b == A & esc\_PSI == A & esc\_RD == B (55:4.264%, 1.0)  
esc\_P == B <= esc\_PSI == C & esc\_SL50 == B (55:4.264%, 0.98)  
esc\_P == B <= esc\_PSI == B (319:24.729%, 0.66)  
esc\_P == C <= esc\_SL50 == C & esc\_EL == Elevado (82:6.357%, 0.76)

esc\_P == B <= esc\_\_b == A & esc\_RD == B & esc\_SL50 == B (43:3.333%, 1.0)  
esc\_P == A <= esc\_\_b == B & esc\_RD == D (73:5.659%, 0.56)  
esc\_P == C <= esc\_\_b == C & esc\_SL50 == C (107:8.295%, 0.64)  
esc\_P == B <= esc\_PSI == A & esc\_RD == B (118:9.147%, 0.77)  
esc\_P == A <= esc\_\_b == C & esc\_RD == D & esc\_SL50 == B (52:4.031%, 0.63)  
esc\_P == C <= esc\_\_b == C & esc\_RD == B & esc\_SL50 == C (57:4.419%, 0.77)  
esc\_P == C <= esc\_PSI == D & esc\_EL == Medio (150:11.628%, 0.57)  
esc\_P == C <= esc\_\_b == C & esc\_RD == B (126:9.767%, 0.6)  
esc\_P == A <= esc\_\_b == B & esc\_PSI == A & esc\_RD == D (36:2.791%, 0.72)  
esc\_P == C <= esc\_PSI == D & esc\_RD == B (129:10.0%, 0.59)  
esc\_P == A <= esc\_\_b == A & esc\_PSI == A & esc\_RD == D (42:3.256%, 0.67)  
esc\_P == C <= esc\_PSI == C & esc\_RD == B & esc\_SL50 == C (30:2.326%, 0.93)  
esc\_P == A <= esc\_\_b == C & esc\_RD == D (78:6.047%, 0.5)  
esc\_P == B <= esc\_\_b == A & esc\_SL50 == B (92:7.132%, 0.78)  
esc\_P == B <= esc\_\_b == A & esc\_RD == B & esc\_SL50 == A (29:2.248%, 1.0)  
esc\_P == B <= esc\_\_b == B & esc\_PSI == B (124:9.612%, 0.71)  
esc\_P == C <= esc\_\_b == B & esc\_RD == B & esc\_SL50 == C (57:4.419%, 0.7)  
esc\_P == C <= esc\_\_b == C & esc\_RD == A (41:3.178%, 0.78)  
esc\_P == C <= esc\_\_b == C & esc\_PSI == D & esc\_RD == B (36:2.791%, 0.81)  
esc\_P == B <= esc\_\_b == C & esc\_RD == C (91:7.054%, 0.75)  
esc\_P == C <= esc\_\_b == C & esc\_PSI == C & esc\_SL50 == D (17:1.318%, 1.0)  
esc\_P == B <= esc\_\_b == A & esc\_PSI == A (112:8.682%, 0.71)

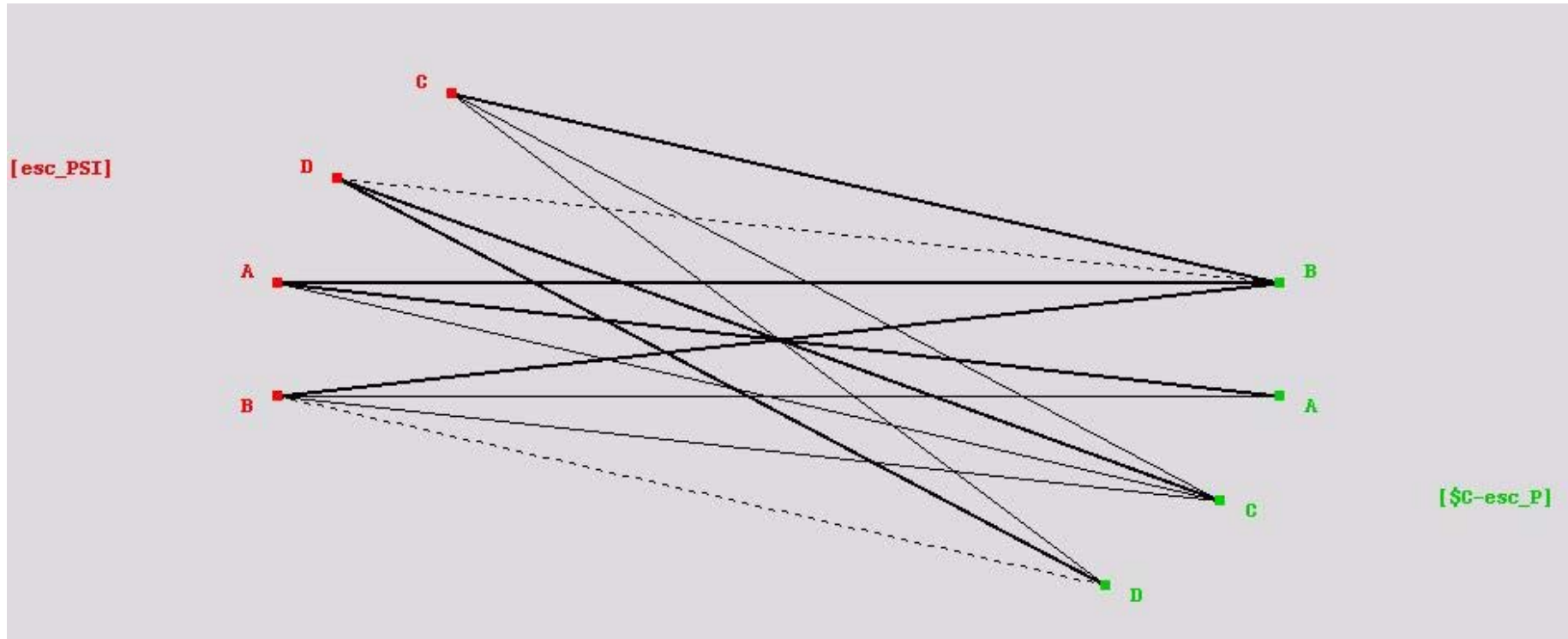
esc\_P == C <= esc\_\_b == B & esc\_PSI == D (60:4.651%, 0.67)  
 esc\_P == B <= esc\_\_b == C & esc\_PSI == C & esc\_SL50 == B  
 (28:2.171%, 0.96)  
 esc\_P == C <= esc\_\_b == C & esc\_PSI == D (71:5.504%, 0.62)  
 esc\_P == B <= esc\_\_b == C & esc\_PSI == C & esc\_RD == C  
 (22:1.705%, 1.0)  
 esc\_P == C <= esc\_\_b == C & esc\_PSI == C & esc\_RD == A  
 (18:1.395%, 0.94)  
 esc\_P == C <= esc\_PSI == C & esc\_RD == B (64:4.961%, 0.62)  
 esc\_P == C <= esc\_RD == C & esc\_SL50 == C & esc\_EL ==  
 Elevado (23:1.783%, 0.87)  
 esc\_P == B <= esc\_\_b == C & esc\_PSI == B (84:6.512%, 0.71)  
 esc\_P == A <= esc\_\_b == A & esc\_RD == D & esc\_SL50 == A  
 (23:1.783%, 0.65)  
 esc\_P == B <= esc\_\_b == D & esc\_PSI == C & esc\_SL50 == B  
 (17:1.318%, 1.0)  
 esc\_P == B <= esc\_\_b == D & esc\_RD == C & esc\_SL50 == B  
 (15:1.163%, 1.0)  
 esc\_P == C <= esc\_PSI == A & esc\_SL50 == C & esc\_EL ==  
 Medio (25:1.938%, 0.76)  
 esc\_P == C <= esc\_PSI == D & esc\_RD == A & esc\_EL ==  
 Elevado (10:0.775%, 1.0)  
 esc\_P == C <= esc\_\_b == C & esc\_PSI == A & esc\_SL50 == C  
 (20:1.55%, 0.8)  
 esc\_P == C <= esc\_\_b == D & esc\_RD == B (22:1.705%, 0.77)  
 esc\_P == D <= esc\_\_b == C & esc\_SL50 == D & esc\_EL ==  
 Muito\_Baixo (8:0.62%, 1.0)  
 esc\_P == D <= esc\_\_b == C & esc\_PSI == D & esc\_EL ==  
 Muito\_Baixo (8:0.62%, 1.0)  
 esc\_P == C <= esc\_PSI == A & esc\_SL50 == C (63:4.884%,  
 0.54)  
 esc\_P == A <= esc\_\_b == B & esc\_PSI == B & esc\_RD == D  
 (30:2.326%, 0.5)

esc\_P == B <= esc\_\_b == D & esc\_PSI == C & esc\_SL50 == A  
 (12:0.93%, 1.0)  
 esc\_P == C <= esc\_PSI == B & esc\_SL50 == C & esc\_EL ==  
 Elevado (30:2.326%, 0.67)  
 esc\_P == C <= esc\_PSI == A & esc\_RD == B & esc\_SL50 == C  
 (26:2.016%, 0.69)  
 esc\_P == C <= esc\_\_b == D & esc\_PSI == D & esc\_EL == Elevado  
 (8:0.62%, 1.0)  
 esc\_P == C <= esc\_SL50 == D & esc\_EL == Elevado (58:4.496%,  
 0.53)  
 esc\_P == B <= esc\_\_b == B & esc\_PSI == A & esc\_EL == Elevado  
 (29:2.248%, 0.79)  
 esc\_P == B <= esc\_\_b == C & esc\_PSI == A & esc\_SL50 == D  
 (10:0.775%, 1.0)  
 esc\_P == A <= esc\_PSI == B & esc\_RD == D & esc\_SL50 == A  
 (24:1.86%, 0.5)  
 esc\_P == D <= esc\_\_b == C & esc\_RD == A & esc\_EL ==  
 Muito\_Baixo (6:0.465%, 1.0)  
 esc\_P == C <= esc\_\_b == A & esc\_SL50 == C & esc\_EL == Elevado  
 (22:1.705%, 0.68)  
 esc\_P == B <= esc\_\_b == D & esc\_RD == C (30:2.326%, 0.77)  
 esc\_P == B <= esc\_\_b == B & esc\_PSI == A & esc\_RD == B  
 (30:2.326%, 0.77)  
 esc\_P == B <= esc\_\_b == D & esc\_PSI == C & esc\_RD == D  
 (9:0.698%, 1.0)  
 esc\_P == C <= esc\_\_b == A & esc\_PSI == D & esc\_EL == Elevado  
 (9:0.698%, 0.89)  
 esc\_P == B <= esc\_\_b == B (299:23.178%, 0.51)  
 esc\_P == C <= esc\_\_b == C & esc\_PSI == A & esc\_RD == B  
 (31:2.403%, 0.58)  
 esc\_P == C <= esc\_\_b == A & esc\_PSI == D & esc\_SL50 == C  
 (37:2.868%, 0.54)

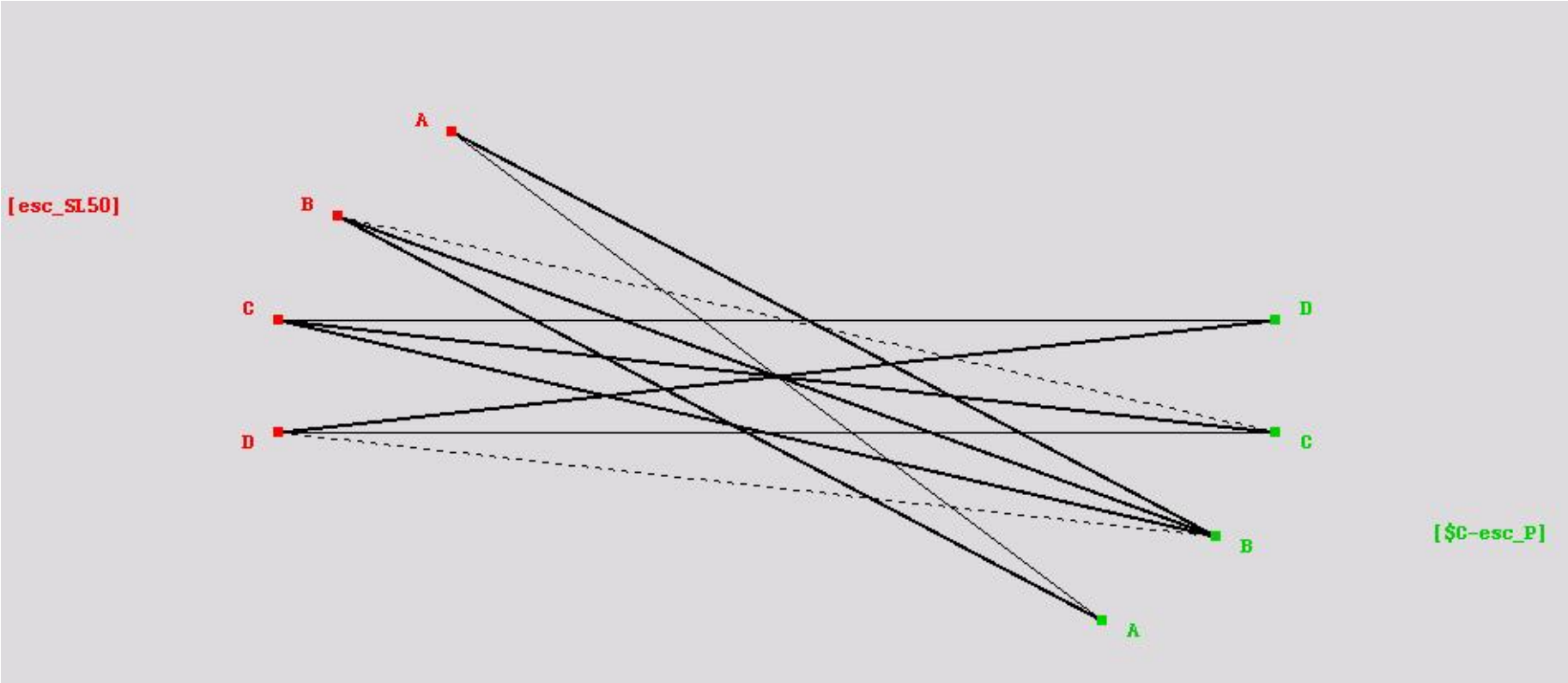
esc\_P == C <= esc\_\_b == B & esc\_PSI == C & esc\_RD == B  
 (16:1.24%, 0.69)  
 esc\_P == C <= esc\_PSI == B & esc\_RD == B & esc\_SL50 == C  
 (46:3.566%, 0.5)  
 esc\_P == D <= esc\_\_b == B & esc\_PSI == D & esc\_EL ==  
 Muito\_Baixo (4:0.31%, 1.0)  
 esc\_P == C <= esc\_RD == A & esc\_SL50 == C & esc\_EL ==  
 Elevado (7:0.543%, 0.86)  
 esc\_P == C <= esc\_\_b == D & esc\_PSI == C & esc\_RD == B  
 (13:1.008%, 0.69)  
 esc\_P == C <= esc\_\_b == D & esc\_PSI == D & esc\_RD == B  
 (4:0.31%, 1.0)  
 esc\_P == C <= esc\_\_b == D & esc\_PSI == B & esc\_SL50 == C  
 (4:0.31%, 1.0)  
 esc\_P == C <= esc\_\_b == A & esc\_PSI == B & esc\_SL50 == D  
 (4:0.31%, 1.0)  
 esc\_P == C <= esc\_PSI == A & esc\_RD == C & esc\_SL50 == C  
 (23:1.783%, 0.57)  
 esc\_P == C <= esc\_\_b == A & esc\_RD == C & esc\_EL ==  
 Elevado (16:1.24%, 0.62)  
 esc\_P == C <= esc\_\_b == C & esc\_PSI == B & esc\_EL ==  
 Elevado (14:1.085%, 0.64)  
 esc\_P == D <= esc\_\_b == C & esc\_EL == Muito\_Baixo  
 (14:1.085%, 0.57)  
 esc\_P == D <= esc\_\_b == B & esc\_RD == B & esc\_EL ==  
 Muito\_Baixo (5:0.388%, 0.8)  
 esc\_P == C <= esc\_PSI == B & esc\_RD == C & esc\_EL ==  
 Elevado (16:1.24%, 0.56)  
 esc\_P == C <= esc\_\_b == A & esc\_SL50 == D & esc\_EL ==  
 Elevado (5:0.388%, 0.8)  
 esc\_P == C <= esc\_RD == C & esc\_SL50 == D & esc\_EL ==  
 Elevado (7:0.543%, 0.71)

esc\_P == D <= esc\_\_b == B & esc\_SL50 == D & esc\_EL ==  
 Muito\_Baixo (2:0.155%, 1.0)  
 esc\_P == C <= esc\_RD == D & esc\_SL50 == C & esc\_EL ==  
 Muito\_Elevado (2:0.155%, 1.0)  
 esc\_P == C <= esc\_\_b == D & esc\_PSI == A & esc\_SL50 == C  
 (2:0.155%, 1.0)  
 esc\_P == C <= esc\_\_b == D & esc\_PSI == A & esc\_RD == B  
 (2:0.155%, 1.0)  
 esc\_P == C <= esc\_\_b == C & esc\_PSI == B & esc\_RD == A  
 (2:0.155%, 1.0)  
 esc\_P == C <= esc\_\_b == A & esc\_RD == D & esc\_EL ==  
 Muito\_Elevado (2:0.155%, 1.0)  
 esc\_P == C <= esc\_RD == A & esc\_SL50 == C & esc\_EL == Baixo  
 (16:1.24%, 0.5)

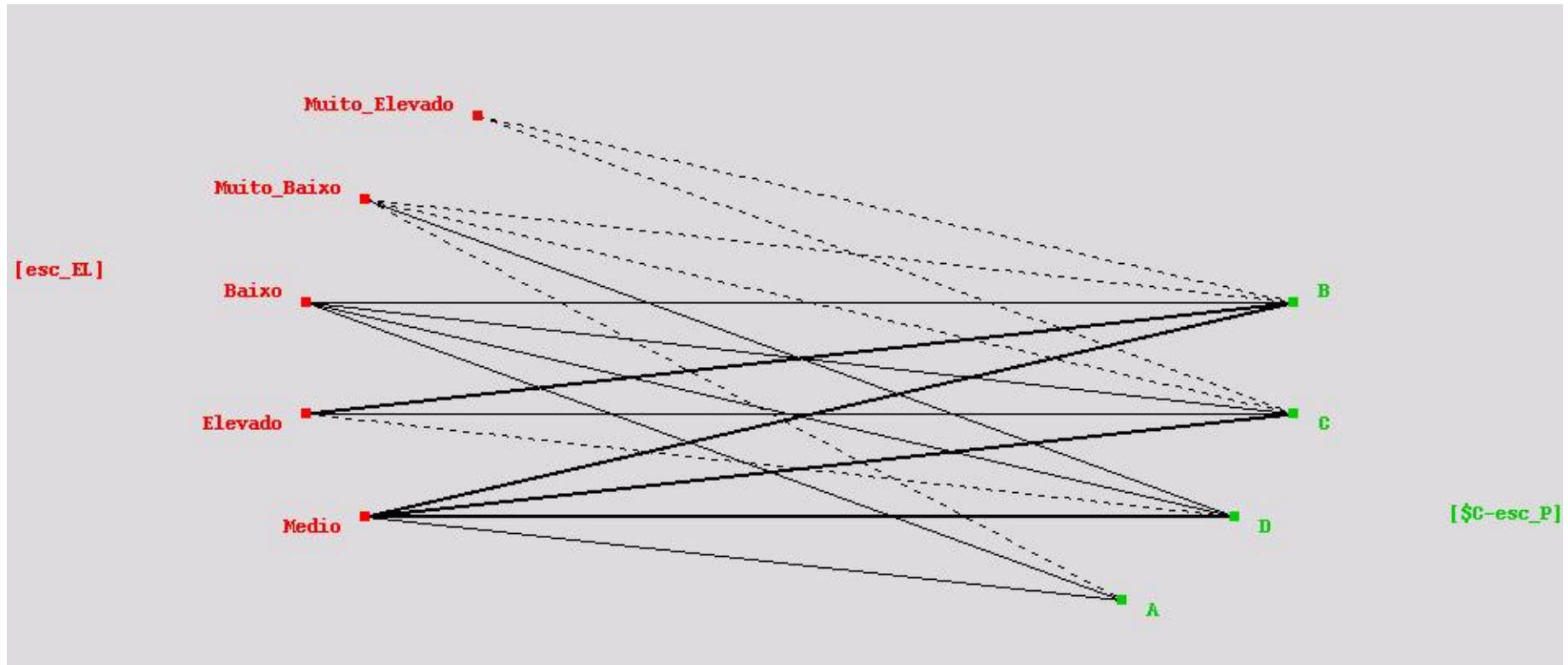
### Esc\_P vs. Esc\_PSI



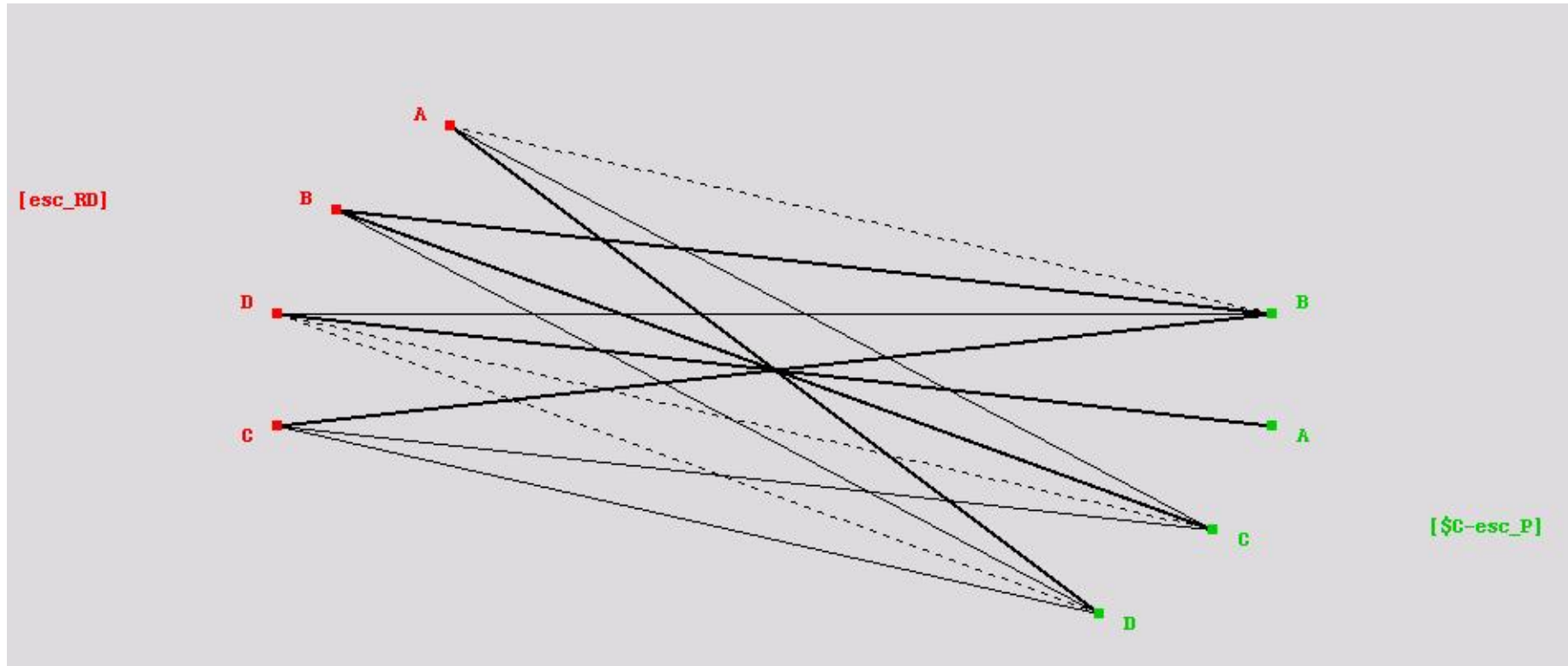
Esc\_P vs. Esc\_SL50



### Esc\_P vs. Esc\_EL



### Esc\_P vs. Esc\_RD



### Esc\_P vs. Esc\_+b

