# 5
# Análise das Conexões

Este capítulo descreve como cenários policiais modelados representados por estruturas em formato de grafo são convertidos em Mapas de Inteligência para análise criminal. Apresentaremos métodos para análise e extração de conhecimento para investigação

O capítulo encontra-se organizado na seguinte seqüência: referencial teórico, método, algoritmos e sistemas utilizados.

## 5.1
## Referencial Teórico

PUC-Rio - Certificação Digital Nº 0412206/CA

Diversas técnicas têm sido aplicadas em análise de redes criminais com objetivo da descoberta de vínculos entre membros de uma quadrilha ou descoberta de conexões entre quadrilhas. A identificação de participantes de agrupamentos criminais pode revelar padrões comuns de procedimentos e fornecer a cronologia das associações criminosas. A análise de vínculos entre entidades pode revelar a existência e a extensão de uma rede criminal, permitindo a identificação de seus limites geográficaos e a modalidade operacional de suas atividades, como organização do narcotráfico e suas possíveis ramificações internacionais.

Wang et al. (2005) citam que criminosos praticam crimes associados através de interesses comuns e recompensas esperadas, como produto dos delitos cometidos. O crime organizado, como terrorismo, narcóticos, seqüestros e roubo a bancos geralmente envolve múltiplos participantes, conectados através de diversos tipos de relacionamentos. Aprender sobre estes interesses e relacionamentos é fundamental para o trabalho de investigação e elucidação de delitos caracterizados por dinâmicas criminais praticadas em cenários complexos.

O emprego do método denominado análise de conexões (link analysis) pode ajudar na descoberta de vínculos existentes em bases de informações policiais, extraindo redes conectadas por objetivos comuns e atividades associadas para cometimento de delitos criminais (Xu & Chen, 2004).

Técnicas de análise de conexões e visualização de redes geradas por informações extraídas de históricos policiais são instrumentos de inteligência na investigação criminal (Goldberg & Senator, 1995).

Segundo Schroeder et al. (2007), a análise de conexões têm sido utilizada do âmbito da justiça em pesquisas de grandes Bancos de Dados para descoberta de entidades associadas, com propósito de facilitar as investigações criminais. Horn et al. (1997) citam como exemplo a ferramenta Link Discovery Tools, uma técnica para análise de conexões utilizada durante a investigação criminal, que produz relacionamentos entre entidades, clusters existentes em sub-redes e identificação de conexões ocultas. Os relacionamentos entre as entidades revelam de que forma as entidades estão associadas na rede (Xu & Chen, 2004).

Segundo Chen & Lynch (1992), em áreas onde são encontrados poucos especialistas, métodos manuais para aquisição de conhecimento não produzem resultados práticos. Por outro lado, sistemas especialistas podem simular a cultura humana através de modelos computacionais, transformando um problema potencial em uma solução especializada residente em um programa de computador.

Algoritmos de caminho mínimo para busca de vínculos fortes entre entidades presentes em uma rede têm sido amplamente usados em diversos domínios temáticos, como pesquisas médicas e redes sociais, não estando devidamente reconhecido como instrumento de análise em ambientes policiais.

Algoritmos de caminho mínimo servem de base para analisar as aproximações entre participantes de um cenário investigado, identificar cumplicidades e estabelecer vínculos fortes entre os atores de uma rede criminal.

Xiang et al. (2005) ressaltam a necessidade da identificação dos diferentes tipos de associações entre criminosos para o melhor entendimento de seus relacionamentos. Devido a dificuldade inerente à análise da atividade criminal a partir de registros textuais, os métodos propostos para descoberta de conhecimento em bases de históricos policiais deveriam incluir:

- Entidades extraídas dos textos e seus relacionamentos para identificação de cumplicidades e autorias
- Mapeamento das entidades e seus relacionamentos com propósito de oferecer melhor compreensão e visualização do conhecimento extraído:

PUC-Rio - Certificação Digital Nº 0412206/CA

- Representações que permitam abstrair com mais facilidade os padrões utilizados pelos atores; facilidade para observação das conexões e topologia geral da estrutura.
- Facilidades para interação com a estrutura através de recursos que permitam a manipulação de dados em diferentes níveis de acesso; visualização das informações através de gráficos sob diferentes formatos da estrutura.
- Recursos para redução do tempo de análise; descoberta de padrões através algoritmos de pesquisa das informações extraídas.

A representação de uma estrutura criminal pode fornecer importantes informações sobre funções desempenhadas por indivíduos em uma rede, particularmente através de seus vínculos com as demais entidades da estrutura. Indivíduos que se revelam muito conectados em uma rede geralmente desempenham funções mais ativas (Horn et al., 1997). Estruturas com diferentes potenciais de atividades, estão representadas pelo volume de conexões diretamente relacionadas com a entidade central da estrutura e com suas extensões e vínculos que se conectam à rede criminal.

Xiang et al. (2005) citam tipos de representação gráfica para relacionamentos (visualizações) e apresentam as vantagens associadas a cada representação:

- Árvore hiperbólica ou estrela;
  - Facilidade para visualização das conexões entre os objetos;
  - Expectativa do usuário melhor reconhecer o conjunto conectado ;
  - Melhor representação de um cluster através de uma árvore hiperbólica;
  - Pode representar um volume maior de objetos na tela.

- Lista em formato de árvore
  - Facilidade para identificação dos objetos conectados;
  - Facilidade para identificação de hierarquias;
  - Similaridade de representação de estruturas com o sistema operacional Windows Explorer.

PUC-Rio - Certificação Digital Nº 0412206/CA

A Figura 5.1 apresenta modelos de representação gráfica de relacionamentos: em formato de lista e em formato Árvore Hiperbólica

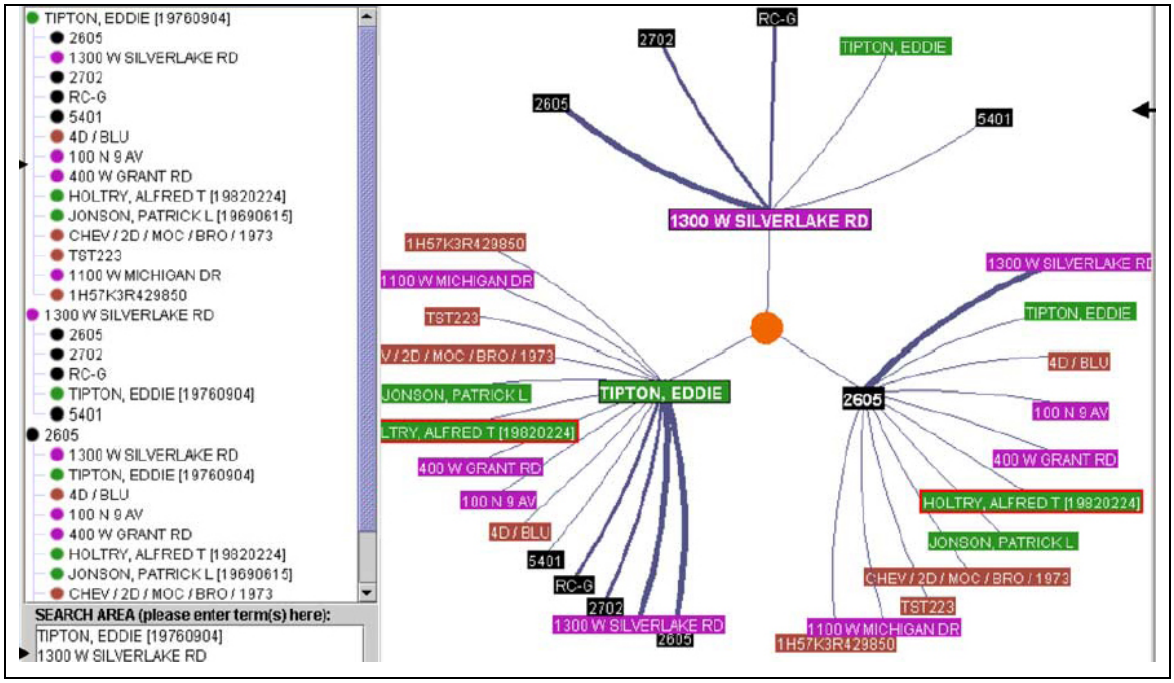



Figura 5.1 – Representação de relacionamentos: formato lista e formato arvore hiperbólica ( Xiang et al., 2005)

### 5.1.1
### Organizações e Estruturas

Diversas estruturas criminais atuantes no Brasil têm adotado comportamento organizacional semelhante a modelos adotados por grupos terroristas integrados por células conectadas por vínculos criminais e ideológicos. Tais estruturas atuam de forma centralizada e descentralizada, cujas atividades dependem do alcance das conexões, como observado em uma cadeia de narcotraficantes.

Segundo Jensen (1999), analistas criminais podem construir Mapas de Inteligência para análise, utilizando uma técnica conhecida como inferência indutiva, que pode prever se uma pessoa poderia, potencialmente, encontrar-se "lavando" dinheiro, baseado em seus depósitos bancários, freqüência das viagens internacionais, conexões de negócios e relacionamentos pessoais. As redes de narcotraficante organizam-se em estruturas conectadas, respondem pela

PUC-Rio - Certificação Digital Nº 0412206/CA

distribuição e suprimento da cadeia de suprimento, contrabando de entorpecentes e lavagem de dinheiro (Xu & Chen , 2004).

Arquilla & Ronfeldt (1996) apresentam na Figura 5.2 três tipos de redes (topologia) organizadas segundo forma de atuação dos criminosos, estruturadas para diferentes propósitos criminais através de tipos distintos de estruturas:

- A estrutura em formato **Cadeia** representa pessoas, objetos ou informações movendo-se ao longo de uma linha de contatos que são separados entre si, e onde um conector de comunicação integra os nós intermediários da seqüência.
- A estrutura em formato **Estrela** ou **Cluster** representa uma estrutura de um **Cartel** onde um grupo de atores estão conectados com um nó (ator) central (não necessariamente hierárquico) e precisam deste nó central para comunicação entre si e coordenação do grupo.
- A estrutura **Multi-Canais** representa uma rede compartilhada por **militantes**, onde todos se conectam com os demais membros do grupo. Esta estrutura requer um eficiente sistema de comunicação, objetivando manter a união entre os componentes da rede. Múltiplos canais fornecem um formato de rede com alto potencial de ação.

PUC-Rio - Certificação Digital Nº 0412206/CA

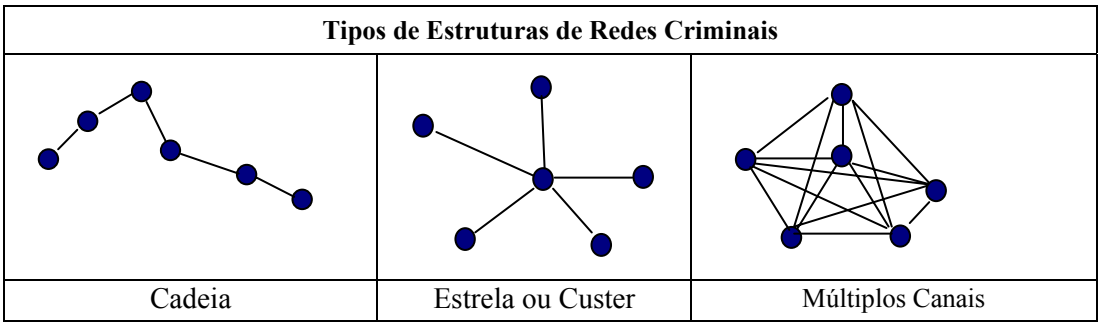



Figura 5.2 – Tipos de estruturas de redes criminais (Arquilla & Ronfeldt ,1996).

Arquilla & Ronfeldt (1996) introduziram o termo "NetWar" para definição de um cenário social emergente, onde criminosos organizam-se, baseiam-se em doutrinas e conceitos, adotam estratégias e comunicam-se, segundo estruturas em forma de rede. Xu & Chen (2008) citam que o conhecimento da estrutura criminal constitui-se em uma das principais chaves para combate a "NetWar". A análise de redes criminais requer a integração de informações de múltiplas incidências criminais e de múltiplas fontes de informações, com o objetivo da

descoberta de padrões sobre estruturas, organizações, operações e do fluxo de informações nas redes criminais. Uma estrutura "NetWar" difere do modo de atuação criminal tradicional, na qual os protagonistas preferem usar estratégias e organizações hierárquicas centralizadas com movimentos maciços de recursos.

Tendo como principal foco de interesse as propriedades estruturais da rede criminal, órgãos de inteligência policial estabelecem como prioridades:

- A identificação das sub-redes que integram a estrutura da rede criminal;
- De que forma os sub-grupos interagem;
- Qual a estrutura da rede criminal;
- Quais são as funções desempenhadas pelos atores nas estruturas criminais.

### 5.1.2
### Aplicações dos princípios de centralidade em Investigações

PUC-Rio - Certificação Digital Nº 0412206/CA

Koelle et al. (2006) citam o uso da análise de redes sociais no combate à ações terroristas e investigações de redes criminais. Wasserman & Faust (1994) citam a importância das conexões entre os atores de uma rede social que representam canais de transferências de recursos. Baker & Faukner (1993) desenvolveram um estudo sobre redes sociais, envolvendo os três indicadores de interação social (*Grau de Conectividade (degree), Intermediação (Betweenness) e Aproximação ou Centralidade (Closeness)* tendo como foco a pesquisa de uma rede clandestina especializada em manipulação de preços no setor de equipamentos elétricos pesados nos Estados Unidos, tais como transformadores e turbinas. Dentre os objetivos principais do estudo desenvolvido por Baker & Faukner (1993) destacam-se a descoberta dos princípios aplicados para invisibilidade da operação ilegal, as funções desempenhadas pelos indivíduos na estrutura clandestina, o cálculo dos indicadores de centralidade da rede e o tamanho da estrutura ilegal.

### 5.1.3
### Identificação de Centralidade de um subgrupo

Uma rede criminal pode ser segmentada em subgrupos integrados por indivíduos que interagem mutuamente. Técnicas de mineração de dados podem detectar e extrair de uma rede grupamentos que não estão claramente visíveis ou claramente identificados. Os resultados obtidos, através de técnicas de *clusterização,* podem identificar indivíduos mais amplamente conectados com os demais membros dos subgrupos, situados hierarquicamente como líderes de subgrupos.

Os subgrupos ou clusters são identificados por padrões ou representação de configurações onde cada elemento do subgrupo apresenta maior similaridade de certa característica ou propriedade com os demais elementos do mesmo agrupamento do que com outros elementos de outros subgrupos. Os subgrupos são geralmente comandados por líderes de quadrilhas especializadas e agrupamentos clandestinos denominados de facções criminais. Os subgrupos desempenham atividades caracterizadas pela tipificação criminal, como seqüestros, roubo de carga, narcotráfico e outros, e distinguem-se pela estrutura organizacional e forma, segundo a qual comunicam-se os seus participantes.

A centralidade de um subgrupo relaciona-se com as funções desempenhadas pelos indivíduos na rede criminal. Segundo Ulrik (2008), centralidade é o conceito mais importante em uma análise de indicadores extraídas de uma rede social. A visão de centralidade é motivada a partir da idéia que uma pessoa mais próxima de outra terá mais acesso à informações que outras pessoas menos próximas (Freeman, 1977). Diversas medidas para centralidade, tais como Grau de conectividade, grau de intermediação ou grau de centralização podem sugerir a importância do nó na rede criminal (Xu & Chen, 2008).

**Grau de conectividade (***Degree***)** - é definido pelo número de conexões do nó com demais nós na rede. Um indivíduo que apresenta um alto grau de conectividade pode estar desempenhando a liderança do grupo.

**Grau de intermediação (***Betweenness***)** - é o número de geodésicos (menor caminho entre dois nós) usando o nó como passagem. Um indivíduo com alto grau de intermediação pode estar funcionalmente operando como gestor de contatos na rede.

PUC-Rio - Certificação Digital Nº 0412206/CA

**Grau de centralização (***Closeness***)** - representado pelo somatório dos menores caminhos entre um nó e os demais nós da rede aos quais tenha acesso.

A comunicação em um grupo social está baseada na propriedade conhecida como grau de intermediação (*Betweenness)*. De acordo com este conceito, a comunicação em uma rede será centralizada em um ponto, se este ponto está situado no mais curto caminho entre dois outros pontos da rede. Um indivíduo localizado estrategicamente em um ponto central particular de uma rede pode influenciar os demais membros participantes da rede através da retenção de informações ou distribuição de informações de forma distorcida ou truncada. A graduação de centralidade é considerada mais forte quando um ponto situa-se no menor caminho entre outros dois pontos do grafo ou situado em seu caminho geodésico de comunicação.

Feeman et. al. (1991) definiram *geodésico* como qualquer caminho conectando dois pontos em um grafo.

No exemplo ilustrado na Figura 5.3 Freeman (1977) apresenta dois pontos geodésicos conectando $p_1$ a $p_3$, o primeiro via $p_2$ e o segundo via $p_4$.

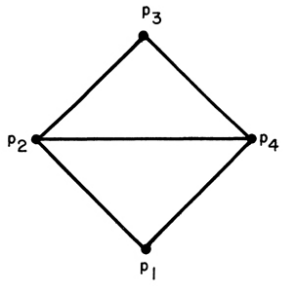



Figura 5.3 – Grafo de cinco arcos de Freeman

Dado um ponto , $p_k$ e um conjunto de pares não orientados $\{p_i , p_j\}$ onde $i \neq j \neq k$ , Freeman (1977) define a proporção de geodésicos que contém $p_k$ com respeito a $(p_i,p_j)$ como:

$$b_{ij}(p_k) = \frac{g_{ij}(p_k)}{g_{ij}}$$

Onde:

$g_{ij}$ = número de geodésicos conectando $p_i$ e $p_j$.

$g_{ij}(p_k)$ = número de geodésicos conectando $p_i$ e $p_j$ que contêm $p_k$

PUC-Rio - Certificação Digital Nº 0412206/CA

$b_{ij}$ ($p_k$) representa a probabilidade do ponto $p_k$ situar-se randomicamente em um goedésico selecionado, conectando $p_i$ e $p_j$.

O potencial do ponto $p_k$ controlar informações entre $p_i$ e $p_j$ pode ser definido com a probabilidade de $p_k$ situar-se em um ponto geodésico entre $p_i$ e $p_j$. Desta forma, a probabilidade de uma mensagem ser transferida através de um ponto geodésico particular é igual a $1/g_{ij}$.

No exemplo ilustrado na Figura 5.4, $p_2$ e $p_4$ apresentam, respectivamente, uma probabilidade igual a ½ de situar-se entre $p_1$ e $p_3$. De forma geral, se $p_k$ situa-se como o único ponto geodésico entre $p_1$ a $p_3$ , então $b_{ij}$ ($p_k$) = 1.

Freeman (1977) apresenta uma fórmula geral para cálculo do índice de centralidade para um determinado ponto $p_k$ em um grafo. Somamos todos os valores parciais referentes à intermediação de $p_k$ para os pares não ordenados, onde i # j # k :

O cálculo processa um somatório geral de todos os valores intermediários (*Betweenness)* em $p_k$ para os pares não direcionados de pontos, onde $i \neq j \neq k$ :

$$C_B(p_k) = \sum_{i<j}^{n} \sum^{n} b_{ij}(p_k)$$

onde:

n = número de pontos no grafo

A soma computada em $C_B$ expressa um índice que representa o grau geral de intermediação de $p_k$.

O ponto máximo de centralidade é obtido apenas quando o número de arcos do grafo é igual a n -1 e existe um ponto $p_k$ situado como geodésico de todos os arcos de valor superior a 1. Este tipo de grafo é denominado como configuração estrela, que representa o valor máximo para $C_B$ , como:

$$\frac{n^2 - 3n + 2}{2}$$

A centralidade relativa de qualquer ponto de um grafo pode ser normalizada, computando-se a expressão, obtida pela divisão do ponto C'b($p_k$) pelo maximo valor possível assumido para $p_k$.

Para este cálculo deveremos considerar o ponto de maior valor associado a **$C'_B(p_k^*)$** dentre todos os pontos investigados na rede. Esta medida é computada

PUC-Rio - Certificação Digital Nº 0412206/CA

pela diferença média entre o mais central dos pontos $C'_B(p_k^*)$ e os demais pontos normalizados do grafo, cujo valor pode variar, de 0 a 1 (Freeman (1977).

$$C'_B(p_k) = \frac{2C_B(p_k)}{n^2 - 3n + 2}$$

Freeman (1977) ilustra alguns exemplos de valores comparados para centralidades relativas, entre diferentes tipos de grafos em estrela (Figura 5.4).

Para qualquer configuração em formato estrela teremos $C'_B(p_k) = 1$ para o nó central, para os demais pontos da rede teremos $C'_B(p_k) = 0$

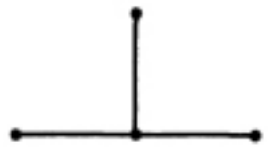
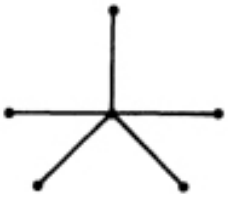
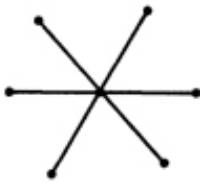
| | | | |
|---|---|---|---|
|  | |  |  |
| N = 3 | N = 4 | N=5 | N=6 |

Figura 5.4 – Exemplos de Centralidades de Grafos em Estrela. Freeman (1977)

Segundo Moody (2003), a centralidade procura identificar quais são os nós de maior importância em uma rede. Intuitivamente, são considerados mais importantes aqueles nós que apresentam maior **Grau de conectividade** ou maior volume de conexões. O conceito que estabelece a importância de um nó relacionado com o seu volume de conexões pode conduzir entretanto a falsas conclusões sobre a estrutura funcional que é representada pelo grafo.

Em termos práticos, a exata identificação de centralidade apresenta propriedades mais complexas para contextualização da importância funcional de um nó em uma rede. Um nó com poucas conexões pode estar situado como "central" em uma rede e desempenhar função de maior importância que outros nós secundários e que apresentam volume maior de conexões.

A Figura 5.5, desenvolvida por Moody (2003), ilustra uma rede onde o maior volume de conexões não relaciona-se diretamente com o nó de maior importância funcional no grafo. O nó [4] apresenta um grau de conectividade = 4 e representa o centro de maior importância funcional do modelo.

PUC-Rio - Certificação Digital Nº 0412206/CA

Os nós rotulados como [6] apresentam um grau de conectividade = 6, estando situados no terceiro nível do escalão funcional do modelo. De acordo com o critério de conectividade, o nó [4] poderia falsamente demonstrar menor importância funcional que os nós [5] e [6].

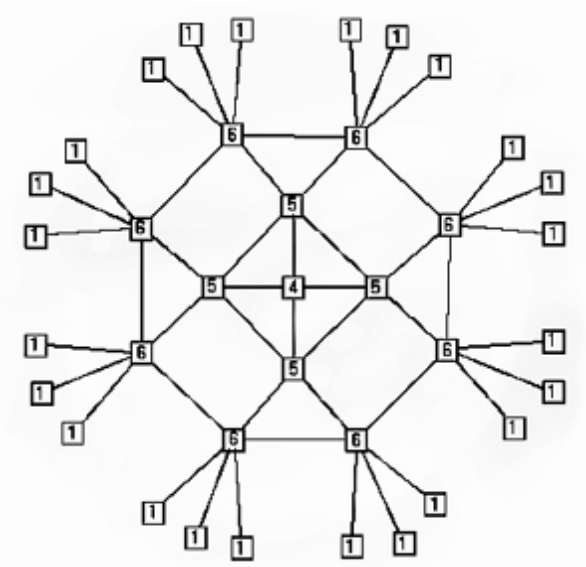


Figura 5.5 – Exemplos de Centralidade não representada por maior grau de conectividade. Moody (2003)

Medidas que definem o grau de centralidade de um grafo baseado na dispersão da centralidade e índice geral de centralização de Freeman foram apresentadas por Moody (2003).

- **Grau de conectividade $C_D$ (*Degree*)** de um ponto $X_i$ do grafo é calculado por:

$$C_D = X_{i+} = \sum_{j}^{n} X_{ij}$$

onde:

n = número de arcos do grafo

$X_{i+}$ arcos que emanam de X

$X_{ij}$ geodésicos do ponto j com passagem em i

- **variância** - calculada pela fórmula seguinte:

$$S_D^2 = \left[ \sum_{i=1}^{g} (C_D(n_i) - \overline{C}_d)^2 \right] / g$$

onde:

g = numero de geodésicos do grafo

PUC-Rio - Certificação Digital Nº 0412206/CA

$C_D$ - numero de geodésico de cada ponto $X_i$

- Indicador de Centralidade geral do grafo, baseado em Freeman - calculado pela fórmula seguinte (Moody, 2003; Carter, 2006; Freeman, 1977):

$$C_D = \frac{\sum_{i=1}^{g}\left[C_D(n^*) - C_D(n_i)\right]}{[(g-1)(g-2)]}$$

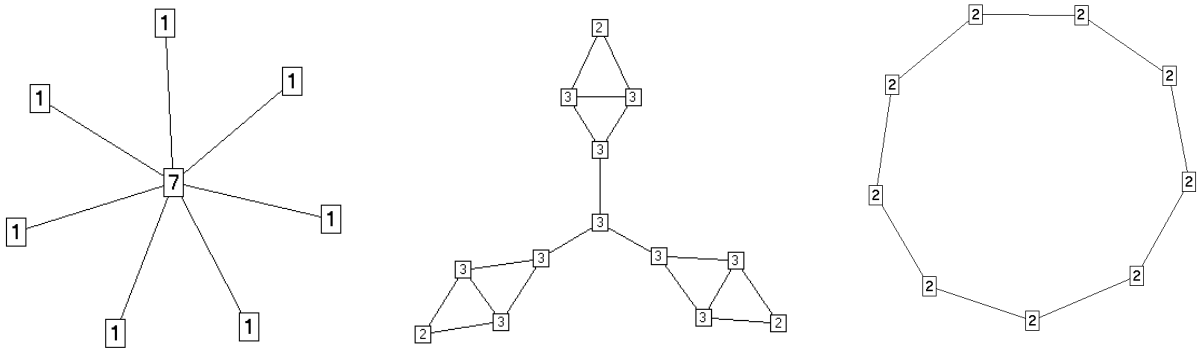
onde:

g = número de geodésicos do grafo

$C_D(n_i)$ = número de geodésicos de cada ponto $X_i$

$C_D$ (n*) = maior dos valores calculados para $C_D(n_i)$

Os exemplos (Figura 5.6) ilustram os indicadores citados por Moody (2003) para calculo de indicadores de centralidade e dispersão.



Ind. Centralidade - 1.0
Variância - 3.9

Ind. Centralidade - 0.2
Variância - 0.17

Ind. Centralidade - 0.0
Variância - 0.0

Figura 5.6 – Indicadores de Centralidade e dispersão em Grafos

Outro exemplo de indicadores aplicados a uma rede mais densa pode ser observado na Figura 5.5, que apresenta uma alta dispersão (variância) e um baixo valor para o Indicador de Centralidade;

- Indicador de Centralidade - 0.1
- Variância - 4.84

O segundo indicador definido por Moody (2003) é a medida de Centralidade ou Aproximação (*Closeness)*. A medida de Centralidade baseia-se no princípio que um ator em uma rede é tão importante quanto a sua maior aproximação relativa de outros atores da rede.

PUC-Rio - Certificação Digital Nº 0412206/CA

A medida de Proximidade é baseada no inverso da distância de cada ator para todos os outros atores participantes da rede, cujo calculo é determinado pela fórmula seguinte:

$$C_c(n_i) = \left[ \sum_{j=1}^{g} d(n_i, n_j) \right]^{-1}$$

Onde $D(n_i,n_j)$ representa a distância entre os nós $(n_i, n_j)$

Medida de Centralidade pode ser normalizada pela fórmula:

$$C'_C(n_i) = (C_C(n_i))(g-1)$$

Onde **g** representa o número de nós do grafo

### 5.1.4
### Centralização em grafos Direcionados

Os indicadores de centralidade apresentados por Moody, 2003) estavam restritos apenas a grafos não direcionados, o que representa uma limitação do potencial de utilidade do método na análise de redes sociais direcionadas. Baseados nos princípios aplicados por Freeman (1979) para cálculo de dependências entre pares de nós para centralidades em redes e na dependência de um nó alcançar qualquer outro nó, usando a mediação de um terceiro nó, White & Borgatti (1994) apresentaram um método para centralidades em grafos direcionados. Este princípio de centralidade que altera o método original de Freeman, aplica caminhos mínimos para definir a relação assimétrica para comunicação entre um par de pontos, quando estes pontos representam pessoas em uma rede social. Um conjunto de arcos conectando dois pontos **i** , **k,** usando **k** como passagem, constitui um caminho de $p_i$ para $p_k$. O caminho mais curto de um ponto para outro ponto é denominado geodésico. Poderá existir mais de um geodésico conectando um par de pontos em um grafo.

Supomos $g_{ik}$ o número de geodésicos de $p_i$ para $p_k$ e $g_{ik}(p_j)$, o número de geodésicos que contém o ponto $p_j$ como nó intermediário nos geodésicos de $p_i$ para $p_k$. A proporção de geodésicos entre $p_i$ e $p_k$ que contêm $p_j$ como ponto intermediário será:

PUC-Rio - Certificação Digital Nº 0412206/CA

$$b_{ik}(p_j) = \frac{g_{ik}(p_j)}{g_{ik}}$$

White & Borgatti (1994) definem ***dependência*** entre pares de pontos em uma rede social como um grau que indica como um certo ponto **p** depende de outro ponto **p** para se comunicar com outros pontos acessíveis em uma rede através de geodésicos. Em uma rede contendo n pontos, teremos como dependência entre os pares $p_i$ e $p_j$:

$$d_{ij}^{*} = \sum_{k=1}^{n} b_{ik}(p_j)(i \neq j \neq k)$$

Podemos calcular a dependência entre quaisquer pares de pontos em uma rede através da matriz:

$$D = (d_{ij}^{*})$$

Cada célula da matriz apresenta o grau de dependência, segundo o qual um ponto da rede, designado por uma linha, depende de outro ponto designado por uma coluna para se comunicar com os demais pontos acessíveis nesta mesma rede.

Como índice genérico de intermediação (Betweenness), aplicado à qualquer ponto de uma rede direcionada, White & Borgatti (1994) desenvolveram o cálculo

$$C_B(p_j) = \sum_{i=1}^{n} d_{ij}^{*}$$

O cálculo desenvolvido por White & Borgatti (1994) foi baseado na fórmula geral de Freeman para grafos não direcionados:

$$\sum_{i=1}^{n} d_{ij}^{*} = 2C_B(p_j)$$

onde cada valor de $d_{ij}$ é duplicado, computado por dois pontos em diagonal, tendo **j** como nó de passagem (Freeman, 1997).

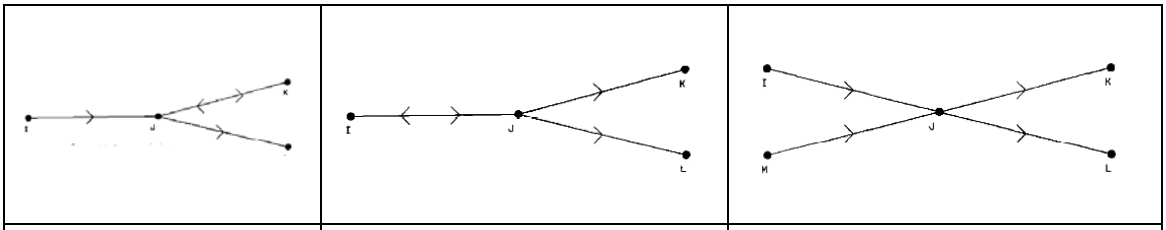
### 5.1.5
### Centralidade Máxima em uma Rede Direcionada

PUC-Rio - Certificação Digital Nº 0412206/CA

Em uma rede direcionada onde, $n_O$ representa o número de pontos com arcos de saída e $n_I$ o número de pontos com arcos de chegada, então em um grafo em formato estrela apresentará o volume de caminhos que passam pelo ponto central igual ao produto $(n_O - 1)\,(n_I - 1)$. Tomemos $n_S$ como o número de pontos com arcos bidirecionais. O mais centralizado dos pontos de um grafo $C_B$ (*Betweenness*) em formato estrela será (White & Borgatti, 1994):

$$C_B(p_j) = (n_1 - 1)(n_O - 1) - (n_S - 1)$$

A Figura 5.7 apresenta exemplos de centralidade em grafos direcionados



| $C_b(p_j) = 3$ | $C_b(p_j) = 2$ | $C_b(p_j) = 4$ |
|---|---|---|

Figura 5.7 – Centralidade em grafos Direcionados (White & Borgatti1994)

PUC-Rio - Certificação Digital Nº 0412206/CA

### 5.1.6
### Descoberta de subgrupos

O problema da descoberta de agrupamentos ou de clusterização pode ser resumido da forma seguinte: supondo um grupo de n itens, este grupo pode ser particionado em subgrupos disjuntos, cuja distância entre os itens é de grande importância. As distâncias mútuas entre todos os itens participantes de um mesmo subgrupo são tão próximas quanto possível, enquanto as distâncias de itens pertencentes a outros subgrupos são tão distantes quanto possível.

Kohonen (1997) apresentou métodos para determinação de subgrupos que denominou de método direto ou interativo e método divisível ou fusão.

O método de divisão segmenta o conjunto inicial sucessivamente até um limite pré-estabelecido quando será então construída uma estrutura em formato de uma árvore mínima expandida que conecta todos os pontos representados através de caminho mínimo. A Figura 5.8 ilustra a divisão de um conjunto em dois subconjuntos e posteriormente a estrutura de pontos resultante em uma árvore expandida mínima.

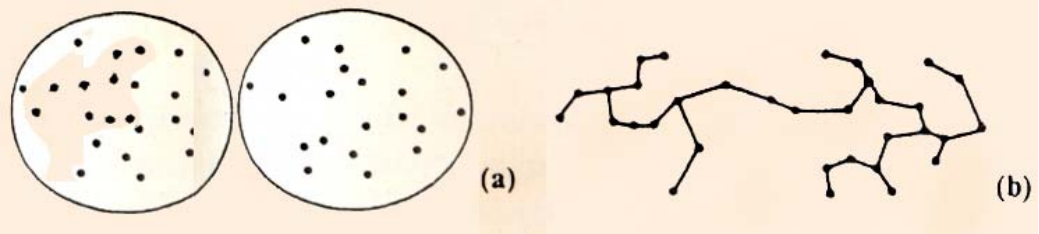



Figura 5.8 – Clusterização através de divisão e árvore expandida (Kohonen, 1997)

Xu & Chen (2008) citam que padrões de interação podem ser identificados usando um algoritmo para análise de redes sociais. Quando aplicado em redes criminais é possível determinar possíveis interações entre subgrupos, detectando a presença ou ausência de associações entre pares de subgrupos, baseada na medida de densidade de conexões.

O método proposto por Xu & Chen (2008), baseado na medida de densidade entre entidades, propõe uma análise de interseções a partir de uma rede segmentada em sub-redes.

Em uma rede representada por arcos não direcionados, a medida de densidade ($\mathbf{D_{ij}}$) entre dois subgrupos **i** e **j** pode ser calculado usando a fórmula a seguir:

$$D_{ij} = \frac{M_{ij}}{N_i N_j}$$

Onde:

$\mathbf{M_{ij}}$ corresponde ao número de conexões entre os sub-grupos **i** e **j**

$\mathbf{N_i N_j}$ correspondem ao numero de nós contidos nos subgrupos **i** e **j**, respectivamente.

É possível identificar a presença de uma associação do tipo *Betweenness* quando a densidade entre dois subgrupos é grande, e cujo valor encontra-se situado acima de determinado limite pré-estabelecido, indicando que os subgrupos interagem com freqüência e possuem uma forte associação (Xu & Chen, 2008).

### 5.1.7
### SOM - Mapas Auto organizáveis

Um método utilizado em redes neurais, conhecido como mapas auto-organizáveis, é utilizado para classificação de informações de entrada, onde encontram-se classes desconhecidas ou padrões organizacionais inerentes aos

PUC-Rio - Certificação Digital Nº 0412206/CA

dados selecionados. Um destes métodos é denominado SOM - Self Organized Maps, que aplica algoritmos de auto-aprendizagem (não supervisionada) onde neurônios competem pela descoberta de padrões de entrada em um método conhecido como ***Competitive Learning*****,** uma forma de aprendizado que particiona o conjunto de padrões de entrada em grupos específicos de dados (Vellasco, 2005).

Segundo Haykin (1994), o objetivo de um algoritmo para aprendizagem auto-organizada é descobrir padrões significativos ou características nos dados de entrada, fazendo esta descoberta em modo não supervisionado. Para desenvolver a aprendizagem, o algoritmo dispõe de um conjunto de regras que o capacita a aprender a calcular um mapeamento de entrada-saída com propriedades desejáveis específicas. O termo local significa que a aplicação ao peso sináptico de um neurônio está limitada à vizinhança imediata daquele neurônio. O processo de aprendizagem consiste em modificar repetidamente os pesos sinápticos das conexões de uma rede neural em resposta a padrões de ativação, atendendo a regras pré-estabelecidas, até se desenvolver uma configuração final.

Técnicas de classificação têm sido usadas para identificação de agrupamentos dotados de características homogêneas, também conhecidos por classes (Kohonen, 1997).

Assumindo o desconhecimento de classes existentes ou relacionamentos entre as entidades de um grupo a priori, Kohonen (1997) cita o problema denominado de *classificação não supervisionada* onde são empregados métodos de clusterização de dados para identificação de subgrupos.

Segundo Haykin (1994), o principal objetivo de um mapa auto-organizável (SOM) é transformar um padrão de entrada de dimensão arbitrária em um mapa uni ou bi-dimensional, convertendo as entradas de forma adaptativa em um mapa topologicamente ordenado. São três os processos essenciais envolvidos na formação de um mapa auto-organizável:

- **Competição** - o neurônio com maior valor da função discriminante, que fornece a base para competição, é declarado vencedor.
- **Cooperação** - o neurônio vencedor determina a localização espacial de uma vizinhança topológica de neurônios excitados, fornecendo a base de cooperação para os neurônios vizinhos.

PUC-Rio - Certificação Digital Nº 0412206/CA

- **Adaptação sináptica** - permite que os neurônios excitados aumentem os seus valores individuais da função discriminante com relação ao padrão de entrada ajustando os seus pesos sinápticos.

**Características da solução Competitive Learning**

Supondo que todos os vetores de entrada e vetores de peso estão normalizados – tem-se todos os vetores dentro de um circulo de raio 1.

No exemplo ilustrado na Figura 5.9 apresentam-se parcialmente vetores de entrada, representados por nomes de indivíduos, distribuídos para uma competição pela busca por padrões de configurações organizacionais. O raio denominado de W4 (neurônio) representa um peso sináptico que competirá com outros pesos existentes pela busca de padrões pré-estabelecidos de entrada.

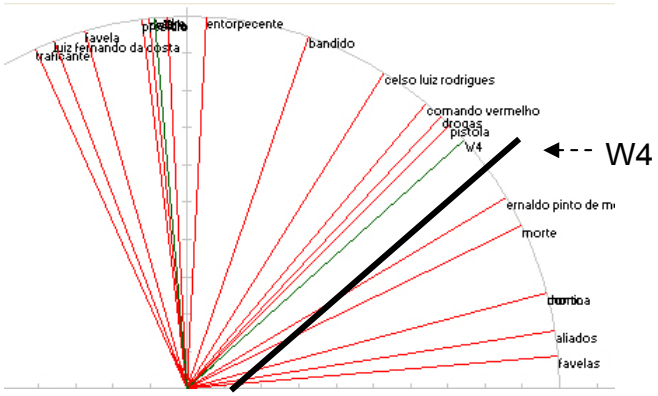



Figura 5.9 – Vista parcial de uma pré-competição SOM

A competição ajusta os pesos de forma que os vetores de entrada similares ativem o mesmo neurônio.

Somente o peso do neurônio vencedor é atualizado porque será o que melhor representa o padrão de entrada do grupo ao qual pertence.

A atualização dos pesos do processador vencedor aproxima ainda mais o vetor **W** do vetor de entrada corrente.

A competição utiliza uma taxa de aprendizado (**α**), normalmente assumindo valor adaptativo entre 0.5 e 0.1, que decresce monotonicamente com o aprendizado.

Kohonen (1997) cita que os dados devem ser normalizados antes da aplicação do algoritmo de auto organização. Este procedimento não é obrigatório, porém recomendável, face o provimento de melhor acuracidade numérica e tratamento dos limites dos dados de entrada.

PUC-Rio - Certificação Digital Nº 0412206/CA

Outras aplicações utilizando redes neurais foram relatadas na bibliografia referenciada para esta Tese. Hauck (2002) cita a aplicação de redes neurais para extração de entidades de textos pela polícia de Tucson, Texas. Xu & Chen (2004) citaram o uso de redes neurais para identificação de caminhos mínimos em grafos. Ali & Kamoun (1993) aplicaram uma rede Hopfield de duas camadas para pesquisa do caminho mínimo em grafos.

### 5.1.8
### Ferramentas existentes para Descoberta de Vínculos

Segundo Xu & Chen (2004), diversas ferramentas propõem-se a executar análise de conexões, entretanto provêm apenas a representação gráfica da rede, não executando análise, dentre estas, destacam-se:

- **Anacapa** - usado extensivamente por departamentos policiais desde o seu lançamento. Esta ferramenta executa a extração e associação de entidades manualmente, que podem receber diferentes símbolos para suas representações visuais. Não dispõe de facilidades para pesquisa de associações exigindo análise manual de vínculos ou caminhos entre entidades. Pode gerar representações gráficas de redes, apresentadas em duas dimensões.
- **Analyst´s Notebook** - desenvolvido para representações gráficas de associações criminais
- **Watson** - permite a pesquisa e identificação automática de associações entre entidades extraídas em um banco de dados. A partir de uma determinada entidade fornecida, Watson procura e fornece outras entidades que possuam alguma referência (vínculo) com a entidade selecionada. A apresentação da associação é executada através de um gráfico.
- **Coplink** - ferramenta para análise de vínculos com capacidade de identificar associações entre entidades e suporte funcional para pesquisas.

PUC-Rio - Certificação Digital Nº 0412206/CA

## 5.2
## Algoritmos utilizados

Neste tópico são apresentados os referenciais teóricos sobre algoritmos pesquisados e técnicas utilizadas para construção de uma matriz de relacionamentos representando um grafo.

### 5.2.1
### Introdução

Algoritmo de caminho mínimo (Xu & Chen, 2004) é um tipo de busca de caminhos que podem identificar o caminho ótimo entre os nós de um grafo, examinando os pesos associados aos seus arcos.

Boaventura (1979) descreve a divisão de algoritmos de caminho mínimo por classes, de acordo com as técnicas de busca utilizadas:

- **Algoritmos por ajustes sucessivos** - trabalham na busca pela minimização do valor temporário de etiquetas associadas aos vértices do grafo até que não seja possível reajustar nenhum valor de comprimento.
- **Métodos por indução e ajuste** - operam com sub grafos progressivamente maiores, até que tenham abrangido todo o grafo; os valores vão sendo minimizados e definitivos até atingir todo o grafo.

Dentro de cada classe distingue-se os algoritmos destinados a procura do caminho mínimo a partir de um vértice dado e os que determinam os caminhos mínimos entre todos os pares de vértices, sendo os mais conhecidos, os seguintes algoritmos ( Boaventura, 1979):

- De ajustes sucessivos
    - Um vértice fixo - Roy, Moore, Dantzig, Dijkstra, Yen, Ford, Bellman-Kalaba
    - Todos os Pares - Bellman-Kalaba generalizado, Farbey, Land e Murchland
- Indução e Ajuste
    - Um vértice fixo - Dantizig-Blattner-Rao
    - Todos os Pares - Dantzig, Floyd, Shimbel, Yen

PUC-Rio - Certificação Digital Nº 0412206/CA

O algoritmo Dijkstra (Dijkstra, 1959) é o método clássico para cálculo do caminho de custo mínimo a partir de um nó fonte para todos os demais nós de um grafo capacitado, pressupondo um grafo contendo arcos não negativos (Xu & Chen, 2008; Boaventura, 1979). Dijkstra é um algoritmo do tipo etiquetado, onde estimativas temporárias são otimizadas e consideradas como permanentes à cada interação.

Dijkstra cedeu os seus fundamentos para os mais eficientes algoritmos voltados para a busca de soluções para o caminho mínimo, cujos princípios foram baseados nas estruturas originais do algoritmo Dijkstra.

Xu & Chen (2004) citam que, em uma rede criminal representada por um grafo direcionado, o valor de uma conexão, que pode assumir um número entre zero e um, pode ser tratado como uma medida da probabilidade para aproximação entre duas entidades diretamente conectadas. Como regra geral, a probabilidade de ocorrência conjunta de um grupo de eventos mutuamente independentes é igual ao produto das probabilidades individuais de ocorrência destes mesmos eventos.

Se dois nós de um grafo se encontrarem apenas através de uma seqüência de conexões intermediárias, o valor da associação entre os dois nós será igual ao produto dos pesos intermediários.

Na Figura 5.10, se o nó A e o nó C são conectados através de B e o peso das conexões intermediárias (A-B) e (B-C) são respectivamente 0,5 e 0,8, desta forma o peso do caminho (A-B-C) deveria ser (0,5 x 0,8) 0,4.

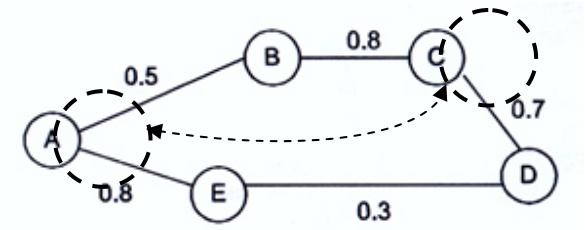



Figura 5.10 – Dois nós indiretamente conectados (A e C) (Xu & Chen, 2004)

Para localizarmos a mais forte associação entre um par de nós deveremos localizar o maior produto dos pesos entre os nós.

Uma vez que os algoritmos de caminho mínimo reconhecem as menores distâncias entre os nós de um grafo, onde o valor dos arcos indica o peso das

PUC-Rio - Certificação Digital Nº 0412206/CA

associações, a representação das conexões mais fortes aplicando algoritmo de caminho mínimo não garante que as mais fortes associações sejam identificadas.

Xu & Chen (2004) propõem uma heurística para transformação da busca pelo caminho mínimo na localização das mais fortes conexões em um grafo direcionado, usando a transformação logarítmica:

$$l = -\ln(w) \qquad 0 < w \leq 1$$

onde:

l é o peso da conexão no novo grafo transformado

w é o peso correspondente no grafo original

Com a transformação proposta obtêm-se os axiomas seguintes:

- Todas as conexões no grafo transformado são números não negativos.

  Como $0 < w \leq 1$, então: $\ln(w) \leq 0$ , logo: $-\ln(w) \geq 0$;

- Os menores valores dos arcos no grafo transformado correspondem aos maiores valores no grafo original.

  Se $l_1 < l_2$ , então $-\ln(w_1) < -\ln(w_2)$ , assim $\ln(w_1) > \ln(w_2)$. Uma vez que ln(w) é uma função monotônica crescente, segue-se que $w_1 > w_2$;

- Os menores caminhos usando a soma dos valores dos pesos no grafo transformado correspondem aos maiores produtos dos arcos usando a rede original.

  Consideremos um caminho mínimo P entre um par de nós A e B onde $l_1, l_2, l_3 \ldots l_n$, representam os pesos associados ao caminho P, sendo $1 \leq p \leq n$, e $n$ o total de nós do grafo. O valor total do caminho P é igual a $\sum_{i=1}^{p} l_i$ . Consideremos outro caminho Q entre A e B, representado pelas conexões $l'_1, l'_2, l'_3 \ldots l'_n$ , sendo $1 \leq q \leq n$. O valor total do caminho Q é igual a $\sum_{i=1}^{q} l'_i$ .

  Por definição, P é o menor caminho entre A e B, então $\sum_{i=1}^{p} l_i < \sum_{i=1}^{q} l'_i$ , e considerando que $l_i = -\ln(w_i)$ e $l'_i = -\ln(w'_i)$ , então concluímos que $\sum_{i=1}^{p} \ln(w_i) < \sum_{i=1}^{q} \ln(w'_i)$

PUC-Rio - Certificação Digital Nº 0412206/CA

Segue-se que $\exp(\sum^{p}_{i=1} \ln(w_i)) > \exp(\sum^{q}_{i=1} \ln(w'_i))$, o que implica em

$$\prod^{p}_{i=1} w_i > \prod^{q}_{i=1} w'_i$$

### 5.2.2
### **Algoritmo PFS Modificado** (Xu & Chen, 2004)

Dentre algoritmos derivados do algoritmo Dijkstra, Xu e Chen (2004) destacam os algoritmos PFS (Priority-First Search) e algoritmo Dijkstra Árvore Dupla. O algoritmo PFS é mais rápido que o algoritmo Dijkstra, emprega uma fila de prioridades em sua implementação e o algoritmo Dijkstra Árvore Dupla interage simultaneamente em duas filas para determinação do caminho mínimo entre os nós de um grafo.

O algoritmo PFS (Apêndice F) é uma variação do algoritmo Dijkstra convencional que busca um caminho mínimo entre os nós de um grafo, mantendo uma árvore de extensão mínima **T** enraizada em um nó fonte **s.** Dado um conjunto de nós **K** pertencentes a um grafo, o algoritmo **PFS** modificado, identifica todos os caminhos mínimos do nó **s** a todos os nós contidos em **K**.

## 5.3
## Método Proposto

Este tópico propõe um método para extração de conhecimento em Mapas de Inteligência construídos a partir da extração de entidades úteis de históricos policiais.

O método é desenvolvido em dois passos, conforme ilustrado na Figura 5.11:

PUC-Rio - Certificação Digital Nº 0412206/CA

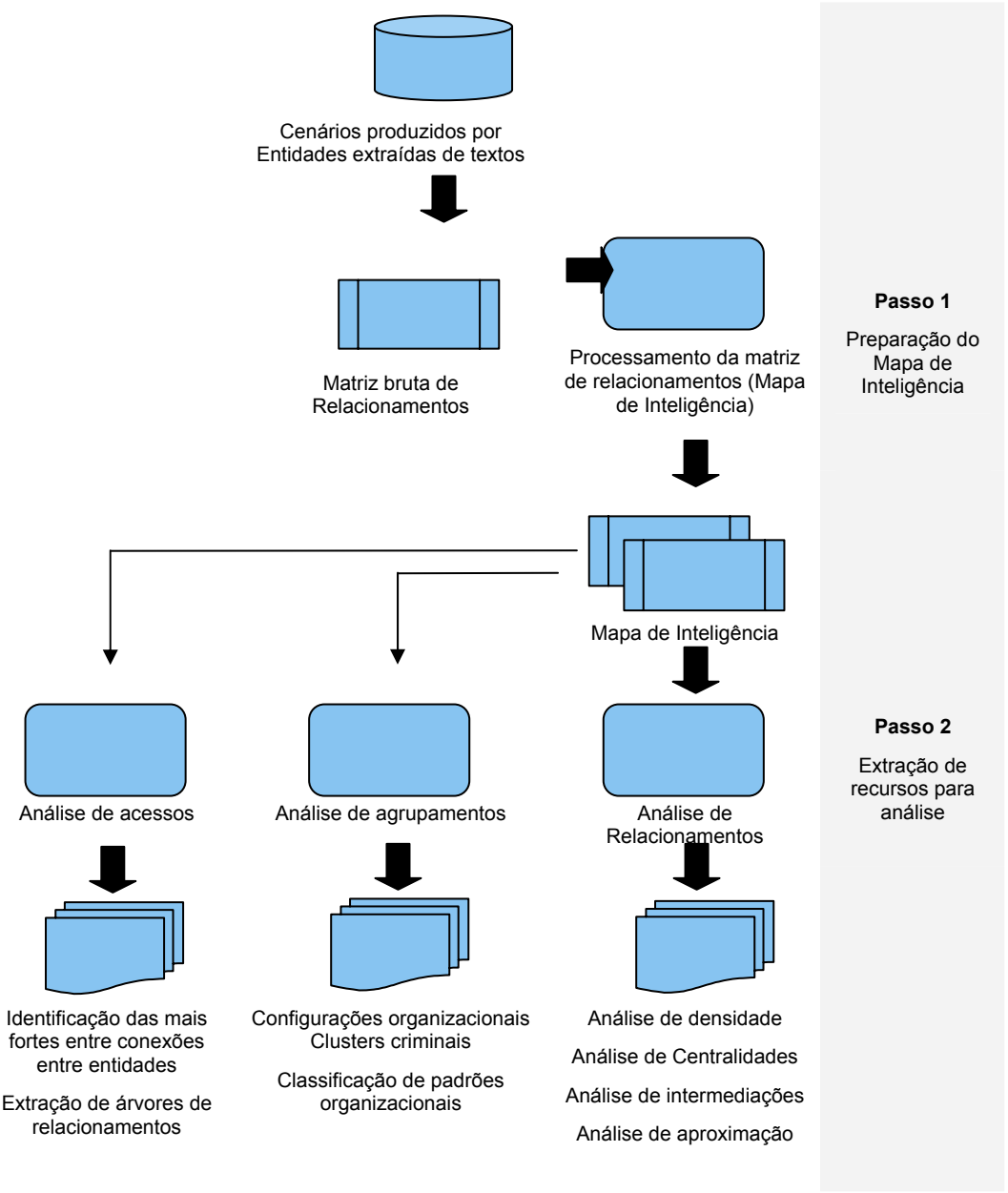



Figura 5.11 – Método proposto para extração de recursos para análise

### 5.3.1
### Preparação do Mapa de Inteligência

Após a modelagem de uma matriz de relacionamentos, representada por um grafo direcionado, construído como produto do relacionamento entre entidades extraídas de arquivos textos, são calculados os mais fortes vínculos entre as entidades do grafo. Os pesos associados aos arcos fornecem cálculo de probabilidades de acesso entre cada par de entidades, apresentando o caminho possível e o valor representativo das mais fortes chances de aproximação entre as

PUC-Rio - Certificação Digital Nº 0412206/CA

entidades. Os resultados calculados são registrados em uma matriz de relacionamentos (Figura 5.12). O cruzamento entre entidades representa os valores calculados (arcos direcionados) para cada associação entre entidades (vértices). A ativação do caminho do mais forte caminho entre entidades usa o algoritmo PFS modificado (Xu & Chen, 2004) que calcula o produtório das probabilidades parciais da trilha de acesso entre quaisquer pares de entidades selecionadas.

A matriz de relacionamentos, que representa o Mapa de Inteligência, e a Matriz de Alcance, são instrumentos contendo informações tratadas que permitem respostas imediatas à investigações, oferecendo respostas imediatas sobre vínculos e valores de aproximação entre entidades, evitando interrupções para o processamento de pesquisas sobre acessos e valores de aproximação entre entidades, ao longo da atividade de análise.

| | luiz fernando d. | comando verm | marcos marinh | marcos antoni | ederson jose g |
|---|---|---|---|---|---|
| luiz fernan | 0 | 100,0% | 81,3% | 43,8% | 13,3% |
| comando | 86,9% | 0 | 71,0% | 38,0% | 12,0% |
| marcos m | 66,0% | 66,0% | 0 | 35,5% | 9,0% |
| marcos an | 31,2% | 31,0% | 31,2% | 0 | 4,0% |
| ederson jo | 8,0% | 8,0% | 7,0% | 4,0% | 0 |
| amigos do | 22,1% | 22,1% | 18,0% | 10,0% | 3,0% |
| celso luiz r | 34,0% | 39,2% | 28,0% | 15,0% | 5,0% |

Figura 5.12 - Matriz de relacionamentos contendo associações pré-calculadas

Como resultado da construção da matriz de relacionamentos é gerada uma matriz que contém trilhas reversas de caminhos possíveis entre pares de entidades presentes no grafo. Esta estrutura denominada Matriz de Alcance, baseada na trilha reversa de acessos do algoritmo Dijkstra, otimiza o uso do Mapa de Inteligência, devido ao pré-cálculo de todos os caminhos possíveis entre entidades vinculadas, evitando o tempo dispendido para processamento das mais fortes associações durante a atividade de investigação.

Cada célula da Matriz de Alcance representa uma referência da entidade identificada pelo número da coluna onde está localizada, indicando os elos do caminho reverso entre os pares de entidades linha/coluna da matriz.

PUC-Rio - Certificação Digital Nº 0412206/CA

### 5.3.2
### Método para extração das mais fortes conexões entre as entidades do grafo

Xu e Chen (2004) propuseram o uso de técnicas para análise de conexões em grafos usando algoritmos de caminho mínimo que são aplicados na identificação das mais fortes conexões entre entidades de uma rede criminal.

O método proposto nesta Tese permite a identificação dos caminhos mais fortes entre entidades, evidenciando os relacionamentos intermediários presentes no caminho selecionado.

Método proposto para esta função

Possibilita investigar acesso e valor da chance de aproximação entre entidades em um Mapa de Inteligência**.** Para isso, na matriz de relacionamentos, seleciona-se um par de entidades quaisquer, considerando-se uma coordenada de referência linha / coluna, a partir de onde deseja-se obter o possível acesso entre as entidades selecionadas e entidades intermediárias porventura existentes entre o par de entidades selecionadas.

Como resposta serão identificados o caminho reverso da mais forte chance de aproximação entre as entidades selecionadas e a probabilidade acumulada de êxito entre as entidades selecionadas

A Extração do mais forte vínculo entre qualquer par de entidades é desenvolvido através de um algoritmo (Apêndice C) que pesquisa uma trilha reversa iniciada com a seleção de uma das entidades identificadas em uma coluna da Matriz de Alcance, retornando sucessivamente as indexações indicadas em um caminho reverso conduzido pela linha de referência do outro par, até que o conteúdo da célula na Matriz de Alcance seja igual a zero (indexação para a própria entidade), finalizando a busca.

Através de uma trilha reversa de caminhos entre entidades o método identifica possíveis vínculos intermediários existentes entre as entidades. As linhas e colunas referenciam os antecedentes de cada passo na trilha de acesso entre entidades. Um exemplo do método pode ser observado no Apêndice G.

PUC-Rio - Certificação Digital Nº 0412206/CA

### 5.3.3
### Método para extração de árvores de relacionamentos

O método permite a representação de uma cadeia de entidades em formato de árvore. A partir de uma entidade qualquer selecionada na matriz de relacionamentos, são extraídos dois níveis de relacionamentos de uma árvore hierárquica. O primeiro nível de relacionamento informa quais entidades estão diretamente vinculadas com a raiz da busca. O segundo nível de relacionamento informa quais entidades subseqüentes estão vinculadas com as entidades do primeiro nível.

Passos da Extração

Uma entidade é selecionada como raiz da busca desejada.

A partir do nó raiz são identificadas todas as entidades vinculadas que emanam do nó selecionado, pesquisados na matriz de alcance.

Para cada entidade vinculada ao nó raiz são identificadas sucessivamentas entidades respectivamente vinculadas na Matriz de Alcance, gerando-se uma árvore com dois níveis hierárquicos.

A estrutura resultante permite a representação de uma árvore de relacionamentos contendo uma entidade raiz e seus vínculos acompanhados, por sua vez, de seus associados.

Os nós internos que estão diretamente conectados ao nó Central são denominados de nós **Centrais** e os nós externos que estão conectados aos nós Centrais são denominados de nós **Satélites**. O nó centralizador da estrutura é denominado de **Principal**.

A Figura 5.13 apresenta um exemplo de árvore de relacionamentos enraizada em uma entidade principal que está representada por uma facção criminal, a partir da qual são identificados vínculos centrais (nível 1) e subseqüentes conexões satélites (nível 2). A representação pode ser gerada sob duas diferentes estruturas: formato lista e formato árvore hierárquica.

PUC-Rio - Certificação Digital Nº 0412206/CA

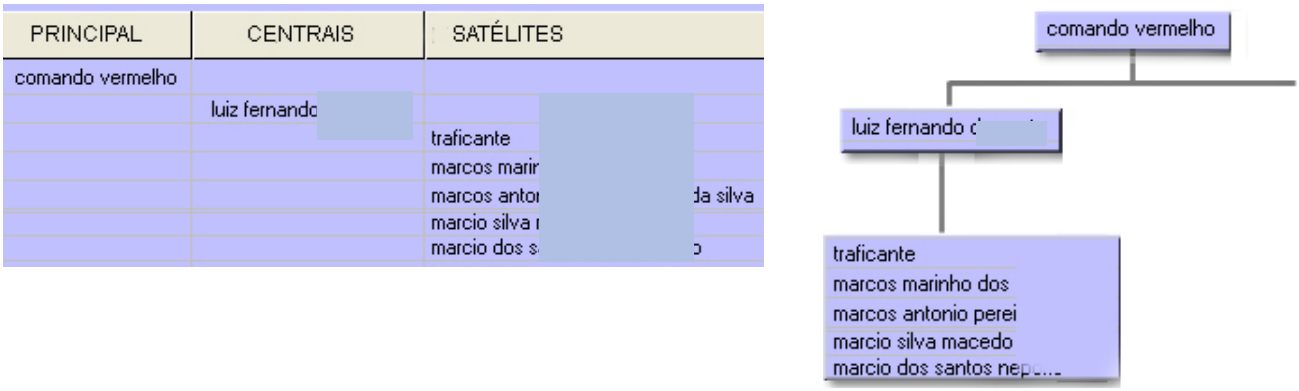



Figura 5.13 – Extração de árvore de relacionamentos

### 5.3.4
### Método para identificação de configurações organizacionais em subgrupos criminais

Subgrupos são caracterizados pelo desenho das conexões associadas à estrutura criminal, através do **Grau de conectividade (degree)** interna do grupo, cujos arcos estão diretamente conectados ao nó principal e do **Grau de conectividade** externa, cujos arcos estão localizados na camada externa do subgrupo. Desta forma é possível extrair para análise os múltiplos formatos organizacionais contidos na estrutura da matriz de relacionamentos do grafo.

Os grupos criminais podem executar as suas atividades através de diferentes modalidades, podendo apresentar estruturas com maior ou menor grau de formalidade operacional, diferenciando a forma de comunicação entre os seus participantes e a facilidade de acesso às lideranças das quadrilhas.

Subgrupos podem ser analisados, de acordo com os vínculos diretos das lideranças (nós centrais) com a chefia da quadrilha (nó principal ou raiz) e subseqüentes conexões satélites (nós satélites) com as respectivas lideranças. Esta análise identifica a forma como se desenvolve a comunicação em uma quadrilha e evidencia traços do seu modo de operação.

#### Descrição do método proposto para esta função

O objetivo deste método é identificar modelos organizacionais em agrupamentos criminais.

PUC-Rio - Certificação Digital Nº 0412206/CA

A partir de um nó raiz no grafo são extraídas cadeias de relacionamentos aplicando-se um algoritmo de busca em profundidade.

As cadeias são extraídas obtendp-se conexões diretamente vinculadas ao nó raiz e conexões diretamente vinculada aos seus sucessores

O resultado apresentado é uma configuração que reflete a organização e hierarquia de um subgrupo criminal.

As seguintes Informações são extraídas através desta função:

- Opção gráfica do modelo organizacional da organização em formato estrela que evidencia o grau de conectividade interior (conectados ao nó raiz) e o grau de conectividade exterior (conectados aos nós centrais).
- Grau de conectividade interior e Grau de conectividade exterior que podem ser caracterizados pelo volume de conexões internas, representadas pelas entidades que estão vinculadas ao nó raiz e pelo volume de conexões externas, representadas pelas entidades que estão vinculadas aos nós diretamente vinculados ao nó raiz

As configurações organizacionais relacionam-se com a especialidade ou com o tipo do crime mais freqüente praticado pelo subgrupo onde encontra-se localizada a organização criminal.

### 5.3.5
### Método para classificação de subgrupos criminais

Subgrupos criminais geralmente apresentam configurações organizacionais semelhantes que podem revelar suas identidades organizacionais.

O estudo de configurações características dos subgrupos extraídos de um Mapa de Inteligência pode revelar importantes informações para análise de redes criminais. Diferentes formatos de conectividade surgem nas configurações dos subgrupos, podendo ser classificados em classes com intuito de se estudar as semelhanças operacionais existentes entre quadrilhas.

Um método é proposto para tratamento de configurações de quadrilhas criminais usando técnicas de classificação não supervisionada de dados através de uma ferramenta que utiliza princípios de um método de redes neurais para

PUC-Rio - Certificação Digital Nº 0412206/CA

tratamento de mapas auto-organizáveis-SOM para aprendizagem não supervisionada.

A função é gerada através da aplicação do Algoritmo AnaphoraSom (Apêndice-E) que trata centróides ou classes agrupadoras de entradas, resultando na segmentação das entradas em estruturas organizacionais semelhantes.

Descrição do Método Proposto para esta função

O objetivo desta função é identificar similaridades organizacionais em subgrupos criminais.

A partir do mapa de relacionamentos são extraídas configurações organizacionais de agrupamentos criminais centralizadas em entidades selecionadas.

Uma função exporta a estrutura contendo a lista de configurações organizacionais para um ambiente de auto aprendizado que aplica princípios de redes neurais para classificação de padrões.

Os dados exportados são normalizados para tratamento posterior pelo Algoritmo AnaphoraSom (Apêndice E).

São atribuídas classes segundo as quais deseja-se agrupar as entradas.

Uma função de ativação **net** de auto-organização (competição) do algoritmo é ativada resultando em agrupamentos que representam as classes que concentram as entradas em estruturas organizacionais semelhantes.

A seguintes Informações são extraídas através desta função:

- Opção de representação gráfica da competição apresentando as entradas distribuídas no hiper-espaço evidenciando os centróides concentradores das entradas (Classes)
- Opção impressa das classes auto-organizadas, por centróides concentradores (classes), onde é informado para cada centróide concentrador (classe) o nome das entradas (entidades) vinculadas após a classificação processada.

### 5.3.6
### Método para identificação de Densidade entre subgrupos criminais

Um método para identificação de densidade entre subgrupos criminais é proposto através de uma função de densidade (Xu & Chen, 2008; Freeman, 1977)

PUC-Rio - Certificação Digital Nº 0412206/CA

que pesquisa a possível interseção existente entre subgrupos criminais identificando elos intermediários (*betweeness*) entre subgrupos selecionados**.**

A **Densidade** é calculada pela relação entre as entidades que funcionam como elos de comunicação entre os participantes dos subgrupos e os demais participantes dos subgrupos selecionados.

Esta função permite visualizar o nível de integração entre dois subgrupos, identificando cumplicidades entre indivíduos e elos comuns entre seus participantes.

Filtros seletivos podem ser aplicados na função de densidade, obtendo-se condições particulares para os resultados extraídos. Os filtros seletivos podem ser combinados de forma restringir a extração à pessoas, logradouros ou entidades temáticas.

PUC-Rio - Certificação Digital Nº 0412206/CA

### Descrição do método proposto para esta função

O objetivo desta função é extrair e medir o nível de comunicação entre subgrupos ou entidades criminais.

Dois subgrupos são selecionados no mapa criminal para análise através de algoritmos de busca em amplitude.

Através do Algoritmo Density (Apêndice-D), uma busca em relacionamentos é aplicada em ambos subgrupos, mantendo as entidades selecionadas como raízes das respectivas extrações.

As entidades que emanam das duas entidades raízes são identificadas extraindo-se configurações em formato estrela para os dois subgrupos.

As entidades comuns entre os subgrupos selecionados são identificadas, computando-se o índice de densidade entre os agrupamentos, calculado por

$$D_{ij} = \frac{M_{ij}}{N_i N_j}$$

onde $M_{ij}$ é igual às conexões comuns entre as lideranças dos subgrupos e $N_iN_j$ é igual ao volume de entidades existentes nos subgrupos.

As seguintes Informações são extraídas;

- **Qualitativas**

  São extraídas as entidades comuns e demais entidades integrantes dos subgrupos selecionados

- **Quantitativa**

  É fornecido um índice de densidade para a interseção entre os subgrupos selecionados

### 5.3.7
### Identificação de Centralidades

O método para identificação de centralidades em subgrupos criminais calcula a medida relativa de aproximação entre cada entidade e as demais entidades de um agrupamento selecionado. O maior dos valores de referência calculados representa a entidade com maior aproximação relativa no interior do subgrupo analisado. O indicador de centralidade também pode ser considerado como uma medida da importância de um indivíduo em um subgrupo (Moddy, 2003; Freeman, 1977). Propomos uma função de centralização que é baseada na computação das maiores chances de êxito (valor de aproximação) entre cada ator e os demais atores em um subgrupo.

#### Descrição do método proposto para esta função

O Objetivo desta função é extrair e medir o nível de centralização $C_c(n_i)$ para cada entidade em um subgrupo criminal.

Um subgrupo criminal é selecionado para cálculo de centralidades.

Cada entidades do subgrupo selecionado é tomado para análise.

Uma árvore é gerada para cada entidade selecionada, mantendo cada entidade como raiz de uma busca em amplitude.

Um índice de aproximação (*Closeness*) é calculado para cada entidade selecionada computando-se as distâncias da entidade para todas as demais entidades da árvore gerada para os arcos direcionados, que são calculados por:

$$C_c(n_i) = \left[ \sum_{j=1}^{g} d(n_i, n_j) \right]$$

onde:

$D(n_i,n_j)$ representa a distância entre os nós $(n_i, n_j)$

O índice de aproximação normalizado é calculado por:

PUC-Rio - Certificação Digital Nº 0412206/CA

$$C'_C(n_i) = (C_C(n_i) / C_C(n^*))$$

onde:

$C_C(n^*)$ - representa o maior valor em $n_i$

g - total de entidades no subgrupo

$d(n_i, n_j)$ - distância entre duas entidades i , j para $i \neq j$

## 5.4
## Sistemas

Os sistemas AnaphoraVisual e VisualSom foram desenvolvidos para processamento dos algoritmos previstos no método de análise proposto por esta Tese.

### 5.4.1
### AnaphoraVisual

O sistema **AnaphoraVisual** foi desenvolvido para prover um ambiente de suporte às funções de análise de conexões extraídas de um Mapa de Inteligência.

O sistema **AnaphoraVisual** utiliza como entrada um ou mais arquivos cenários, tratando as variáveis lidas como elementos de um grafo. As variáveis recuperadas dos arquivos cenários são modeladas em uma estrutura matricial que serve de base para o processamento dos algoritmos de análise criminal e suas funções complementares de relacionamentos entre entidades.

A Figura 5.14 apresenta o painel principal de funções do sistema **AnaphoraVisual**, exemplificando um resultado tratado de um arquivo cenário. As tabelas exibem os nós do grafo, e cada arco ponderado associando os pares vinculados de entidades presentes no grafo lido**.**

PUC-Rio - Certificação Digital Nº 0412206/CA

Figura 5.14 – Painel principal do Sistema AnaphoraVisual

O sistema AnaphoraVisual executa o tratamento de um arquivo cenário produzindo um grafo a partir de entidades extraídas de textos.

A leitura é realizada após seleção de um **arquivo cenário,** a partir de uma caixa de diálogos ofertada pelo sistema. O tratamento do arquivo é executado pelo sistema que converte um arquivo cenário em um Mapa de Inteligência.

Através do processamento dos **arquivos cenários,** matrizes de suporte às funções de análise são produzidas.

As funções seguintes são suportadas pelo sistema AnaphoraVisual.

- **Mapa de Inteligência** - valores pré-calculados contendo as mais fortes conexões entre cada pares de entidades onde o acesso foi possível.
- **Matriz de Alcance** - referências das trilhas reversas para as mais fortes conexões (antecessores dos caminhos de acesso entre entidades).
- **Identificação das mais fortes conexões** - função para geração de uma trilha das mais fortes conexões entre duas entidades selecionadas na matriz de inteligência.
- **Árvore de relacionamentos -** função para geração de uma árvore de conexões. A partir de uma entidade selecionada da matriz de distâncias é gerada uma árvore de relacionamento em formato tabela, listando as cadeias de entidades vinculadas como nós centrais (internos) e entidades conectadas como nós satélites (externos).
- **Cálculo de densidade entre sub grupos -** função para identificação de interseções entre pares de agrupamentos criminais selecionados em um Mapa de Inteligência. A rotina calcula a densidade entre os subgrupos,

PUC-Rio - Certificação Digital Nº 0412206/CA

identifica as entidades existentes nos agrupamentos e analisa os vínculos existentes em sua interseção.

- **Cálculo de centralidade em um subgrupo -** função para identificação as maiores aproximações existentes para cada entidade pertencente a um subgrupo criminal. A rotina calcula um índice relativo de aproximações entre participantes de um subgrupos computando os valores existentes das distâncias entre entidades.
- **Identificação de Vínculos entre sub grupos -** função para identificação quantitativa de vínculos, porventura existentes, entre um subgrupo selecionado e demais subgrupos criminais presentes em uma lista.
- **Exportar Subgrupos para identificação de padrões -** função para exportação de uma estrutura de dados normalizados referentes à configurações contendo quantitativos de conexões centrais e satélites de agrupamentos criminais para tratamento de padrões através de um sistema de mapas auto-organizáveis SOM

A programação para tratamento das funções do sistema **AnaphoraVisual** foi desenvolvida pelo autor em 2007/2008, para os laboratórios requeridos pela Tese, utilizando linguagem de programação Microsoft Visual Basic 6, com extensões para leitura de documentos em formato texto livre (.txt), texto rico (.rtf) e textos formatados em Microsoft Word (.doc).

PUC-Rio - Certificação Digital Nº 0412206/CA

### 5.4.2
### Visual SOM

**VisualSOM** é um sistema para classificação não supervisionada de variáveis que fornece como resultado um mapa auto-organizado SOM através do algoritmo **AnaphoraSOM**.

Os objetivos principais do sistema **VisualSOM** são os seguintes :

- Demonstrar graficamente os resultados promovidos por uma competição entre os processadores de uma rede neural auto-organizável.
- Desenvolver ensaios para um método denominado Competitive Learning que pode ser aplicado como solução para clusterização de dados.

## Representação das entradas

**VisualSOM** é um laboratório, orientado para aplicações de redes neurais, permite ensaios com variáveis normalizadas, agrupando os resultados em classes ou clusters. As variáveis são denominadas "Inputs" e os neurônios agrupadores são denominados "Pesos". Estas informações permanecem residentes temporariamente em uma tabela localizada em um painel de trabalho do programa

Após a leitura das entradas, o sistema VisualSOM distribui as variáveis no hiper espaço, em um circulo de raio 1, onde terá início uma competição do auto-aprendizado. Os dados de entrada alimentados para competição são organizados em uma tabela e convertidos para referências trigonométricas.

A Figura 5.15 ilustra o painel gráfico do sistema VisualSom onde são distribuídas as entradas para classificação. Uma das principais características do sistema VisualSom é apresentar dinamicamente a evolução da competição, permitindo acompanhar a trilha de deslocamento de cada neurônio durante a competição. As entradas opcionalmente podem ser apresentadas no painel gráfico rotuladas com códigos mnemônicos ou com títulos originais de entrada (nomes originais das entradas). Linhas coloridas diferenciadas apresentam o curso de ação de cada neurônio durante a competição. Cada fase da competição utiliza uma taxa de aprendizado que é utilizada pelo algoritmo SOM, que decai monotonicamente em sucessivas iterações, até um limite pré-estabelecido, finalizando o ciclo de aprendizado.

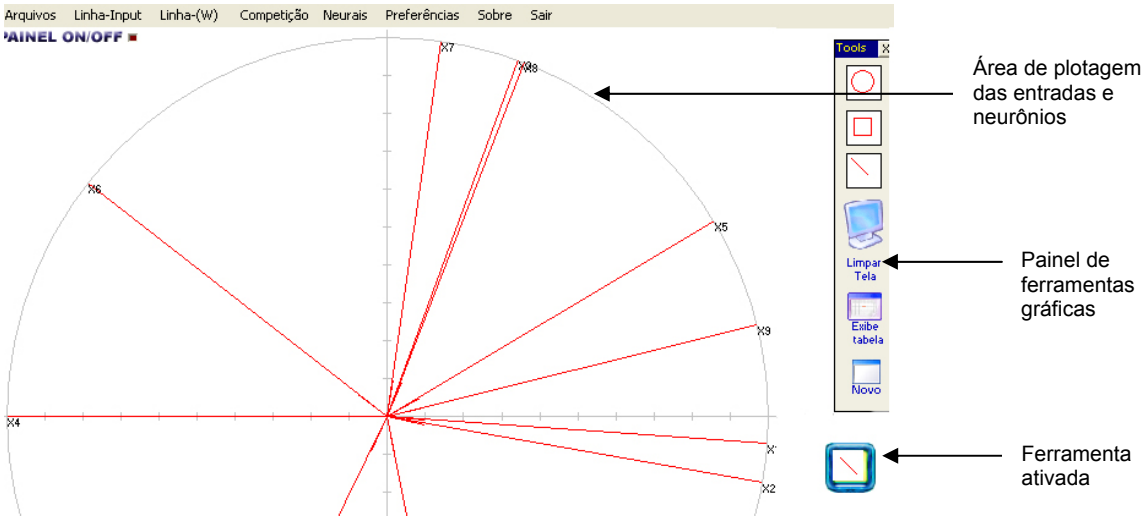



Figura 5.15 – Painel Gráfico de sistema Visual SOM

PUC-Rio - Certificação Digital Nº 0412206/CA

A Figura 5.16 apresenta o painel principal para ativação de funções suportadas pelo sistema VisualSOM.

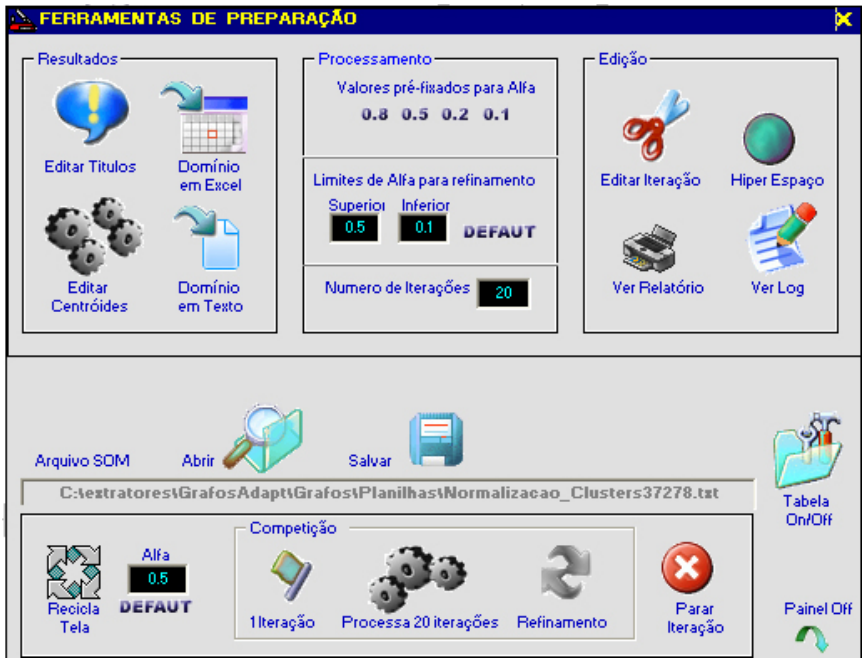



Figura 5.16 – Painel de Funções Visual SOM

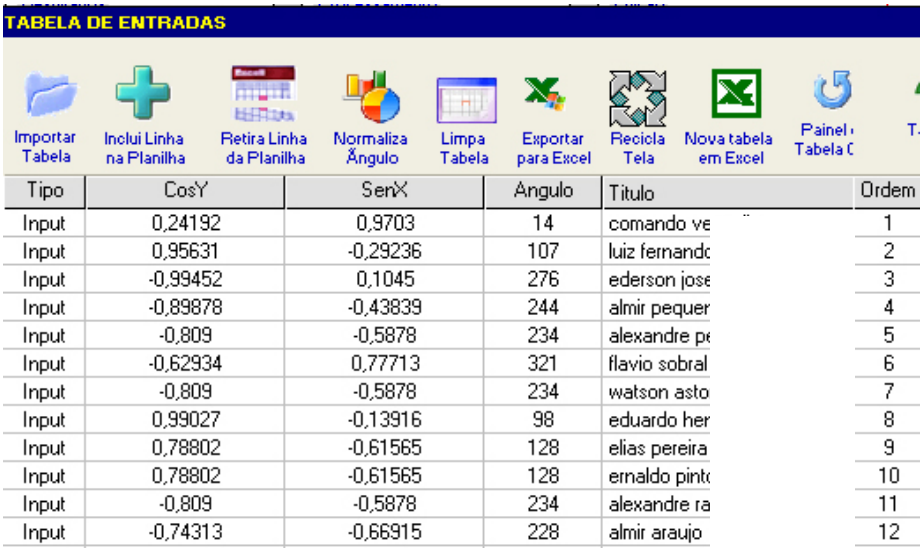
Princípios do modelo processador neural **Competitive Learning**:

O sistema VisualSom utiliza princípios do aprendizado competitivo, que busca padrões organizacionais semelhantes nos dados de entrada. O processamento é realizado a partir de um arquivo temporário exportado pelo sistema AnaphoraVisual, contendo dados normalizados de configurações de subgrupos criminais sobre os quais deseja-se conhecer semelhanças organizacionais. As principais características da função neural SOM são as seguintes:

- Função de ativação - Degrau (apenas para o neurônio vencedor).
- Topologia - Uma única camada de processadores.
- Algoritmo de aprendizado - Não supervisionado.
- Valores de entrada/saída - Binários/contínuos.

A Figura 5.17 apresenta um exemplo de um conjunto de entradas ajustadas para etapa da competição.

PUC-Rio - Certificação Digital Nº 0412206/CA




| Tipo | CosY | SenX | Angulo | Titulo | Ordem |
|---|---|---|---|---|---|
| Input | 0,24192 | 0,9703 | 14 | comando ve | 1 |
| Input | 0,95631 | -0,29236 | 107 | luiz fernando | 2 |
| Input | -0,99452 | 0,1045 | 276 | ederson jose | 3 |
| Input | -0,89878 | -0,43839 | 244 | almir pequer | 4 |
| Input | -0,809 | -0,5878 | 234 | alexandre pe | 5 |
| Input | -0,62934 | 0,77713 | 321 | flavio sobral | 6 |
| Input | -0,809 | -0,5878 | 234 | watson asto | 7 |
| Input | 0,99027 | -0,13916 | 98 | eduardo her | 8 |
| Input | 0,78802 | -0,61565 | 128 | elias pereira | 9 |
| Input | 0,78802 | -0,61565 | 128 | ernaldo pinto | 10 |
| Input | -0,809 | -0,5878 | 234 | alexandre ra | 11 |
| Input | -0,74313 | -0,66915 | 228 | almir araujo | 12 |

Figura 5.17 – Vetores Normalizados Visual SOM

As funções seguintes são suportadas pelo sistema **VisualSOM.**

- **Gerar Competição -** Função destinada à gerar ciclos de classificação das variáveis de entrada em classes pré-especificadas. O sistema realiza n ciclos de competição orientados pelo valor de uma de taxa de aprendizado decrescente em número de iterações previstas para encerramento.
- **Geração de resultados para competição** - os resultados do sistema de competição podem ser apresentados nos formatos seguintes:
  - **Formato Gráfico** - apresentação gráfica dos vetores de entrada e neuronios vencedores da competição, distribuídos em regiões resultantes do processamento de auto-organização.
  - **Formato texto** - As entradas são apresentadas em classes, organizadas segundo os resultados dos clusters gerados pelo algoritmo SOM. A Tabela 5.1 apresenta um exemplo de um resultado de uma competição formato texto.
- Resumo geral das operações de Entrada e Saída
  - Edição na tela os Títulos das Entradas (inputs)
  - Edição na Tela as coordenadas das entradas e centróides
  - Exportação das saídas processadas para uma planilha Excel
  - Exporção das saídas para um arquivo em formato texto
  - Importação de um arquivo contendo as entradas de uma tabela em formato texto

PUC-Rio - Certificação Digital Nº 0412206/CA

- Importação de um arquivo contendo as entradas de uma tabela em formato Excel
- Edita no relatório a imagem do hiper espaço e resumo das propriedades dos neurônios
- Edita no relatório, a imagem gráfica do hiper espaço

Tabela 5.1 – Tabela de Vetores classificados saída texto

| Centróide : 1 | Centróide : 2 | Centróide : 3 | Centróide : 4 |
|---|---|---|---|
| evilao feitosa<br>andn de souza<br>luiz fernndo | marcos dos ss<br>marcos ilva<br>marcio neceno<br>alexsandro galoda silva<br>ede jose leite<br>amigos dos amigos<br>marin dos santos<br>marcos gom roigues | comando vermelho<br>drogas<br>celso luiz Rigues<br>ernao de meiros | pauldos santos<br>trafico<br>armas |

Fonte: O Autor

PUC-Rio - Certificação Digital Nº 0412206/CA

A programação do algoritmo para processamento das competições neurais SOM executadas pelo Sistema **VisualSom** é inédita, tendo sido desenvolvido pelo autor em 2005/2006, para os testes laboratoriais requeridos pela Tese, utilizando linguagem de programação Microsoft Visual Basic 6, com extensões para leitura de planilhas Excel, incorporando para este fim recursos suportados pelo objeto Microsoft OLE DB ODBC Driver.

## 6 - Resumo

Neste capítulo apresentamos como cenários policiais modelados a partir de entidades extraídas de históricos policiais foram convertidos em Mapas de Inteligência.

Foram apresentadas ferramentas para descoberta dos mais fortes vínculos entre entidades e subgrupos criminais extraídos do Mapa de Inteligência, baseadas em algoritmos de caminho mínimo e busca em grafos.

Apresentamos a aplicação de indicadores sociais de centralidade, conectividade e intermediação para identificação de lideranças e posições mais relevantes de comunicação em um subgrupo criminal.

Demonstramos a utilização de um método que aplica redes neurais para classificação de padrões organizacionais de redes criminais baseado no método de mapas auto-organizáveis (SOM).

PUC-Rio - Certificação Digital Nº 0412206/CA