PUC RIO
DEPARTAMENTO DE ENGENHARIA INDUSTRIAL

EXAME DE QUALIFICAÇÃO

Profa. Dra. Marley Velasco
Prof. Dr. José Eugenio Leal
Prof. Dr.Sílvio Hamacher

ORIENTADOR – SILVIO HAMACHER
DOUTORANDO – ISNARD T. MARTINS

Artigo Selecionado:

Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks

Publicado em: **Decision Support System**
Decision Support System 38 (2004) 473-487

Jennifer J. Xu & Hsinchun Chen,
Universidade do Arizona, Tucson, USA, 2003

Síntese do Problema

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

Construção das Redes Utilizadas para Pesquisa

Algoritmos propostos

Conceitos, Algoritmos, Indicadores, Avaliação

Recomendações dos autores

Apreciação

Tema da Tese

(Chen, Coplink, 2004)

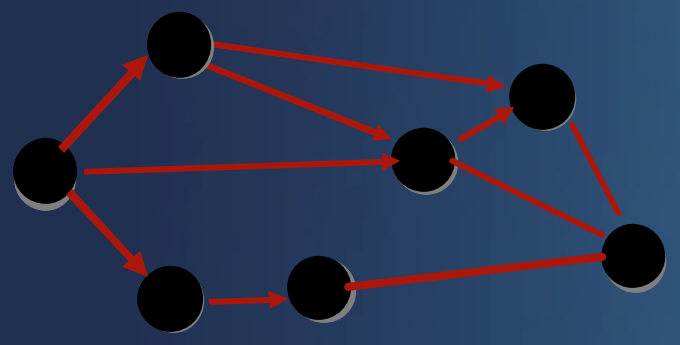
A Síntese do Problema (1) :

Investigações policias utilizam mapas integrando peças do cenário criminal.

Os mapas devem apresentar possiveis conexões entre os atores potencialmente envolvidos

Este mapa, ou árvore de relacionamento criminal pode ser, muitas vezes, tratado como uma rede

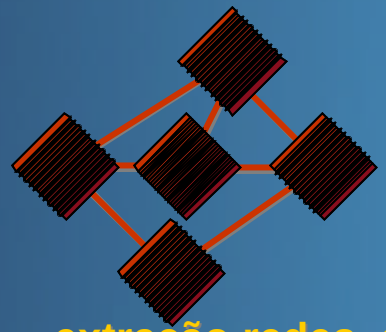
[McAndrew,1999]



A Síntese do Problema (2) :

Históricos policiais

ricos em informações



extração redes de relacionamento

evidências preciosas para investigações

difícil análise

Ocultam conexões entre fatos , objetos, pessoas…

Pesquisas profundas exigem tempo e pessoal qualificado para análise histórica - técnica conhecida como

Link Analysis

…consiste de :

- extração de conceitos ;
- construção da rede;
- identificação e análise de conexões entre as entidades

[Chau, Lynch, 2002] b

A Síntese do Problema (3) :

Entretanto, os mapas de Redes Criminais

…complexos e demandam tempo dos analistas.

…apenas poucos mapas são desenvolvidos

[McAndrew,1999]

Sequestros, homicídios e crimes hediondos,

…o tempo torna-se crítico

Quando envolvem denúncias, múltiplos participantes, lugares, instrumentos, cúmplices e vítimas

…a rede torna-se ainda mais complexa.

Solução?

A Síntese do Problema (4) :

- Ferramentas para preparação automática de redes criminais …são utilizadas

  [Goldberg & Wong , 1998] [W.Harper et al, 1998] [R.Hauk at al 2002]

- Significativo volume de conceitos inúteis são extraídos de forma automática

…dificultam a identificação dos conceitos essenciais à investigação

…dificultam a identificação das fortes conexões entre conceitos relevantes

[P. Klerks et al 2002]

Proposição dos Autores

- **Resumo da proposição**
  - Objetivos da Proposta
  - Desenvolvimento da Pesquisa
    - Revisão de Conceitos
    - Construção das Redes Utilizadas para Pesquisa
    - Algoritmos propostos
      - Conceitos, Algoritmos, Indicadores, Avaliação
  - **Recomendações dos autores**
  - **Apreciação**
  - **Tema da Tese**

(Chen, Coplink, 2004)

Resumo da Proposição dos Autores

Emprego de técnicas de pesquisa operacional para análise de associações existentes em redes criminais;

Avaliação de algoritmos de Caminho mínimo para análise das mais fortes conexões em redes criminais.

- **Um exemplo** [V Krebs, 2001]

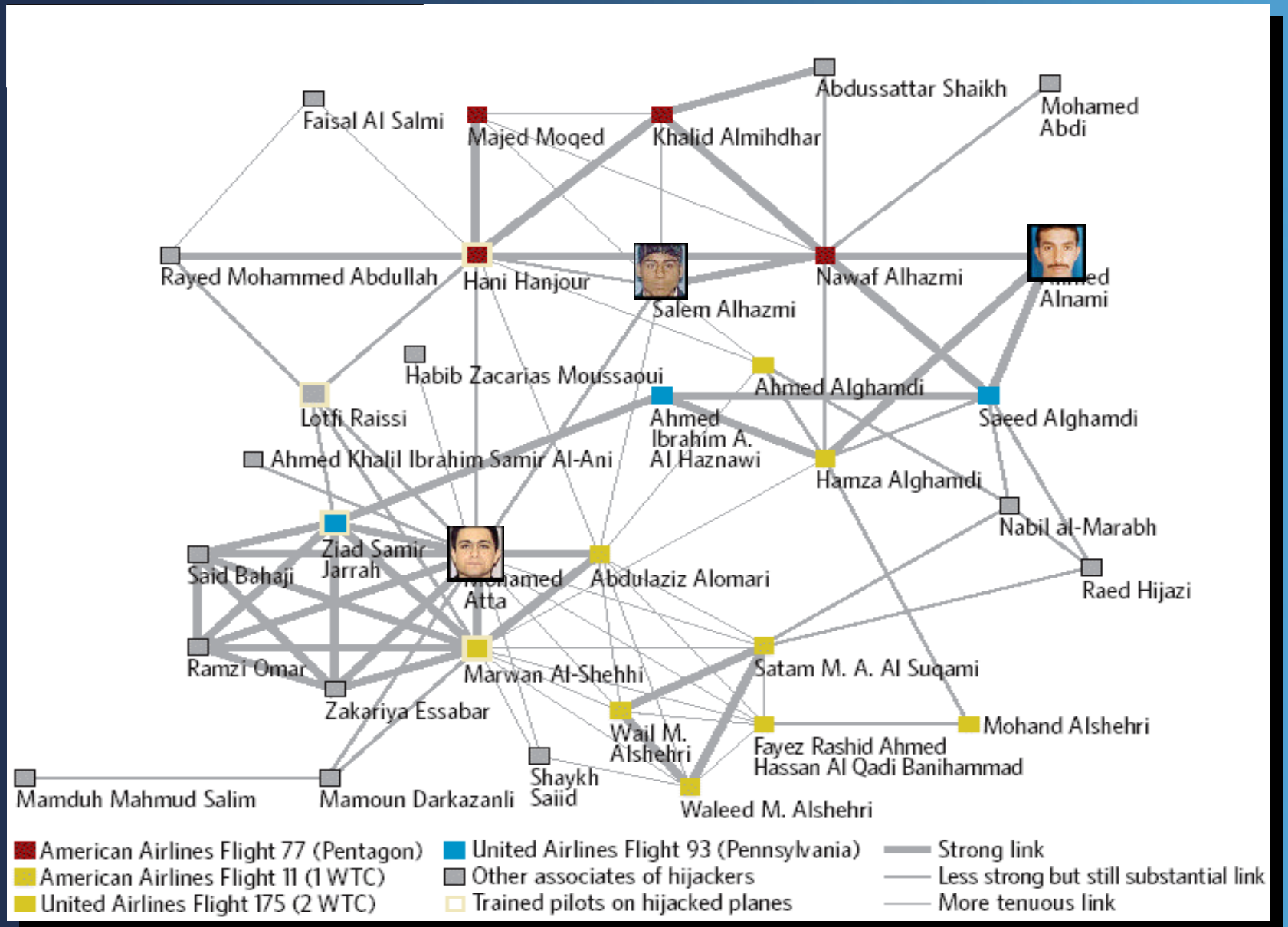
1- Como
   - Mohamed Atta,
   - Ahmed Al Haznawi,
   - e Salem Al Hazmi

…estão relacionados nas ococorrências policiais?

2- Quais são as mais fortes associações entre eles ?

Um exemplo de associações criminosas
A rede terrorista 11 de setembro – EEUU

[Chau, Xu, Chen, 2002]

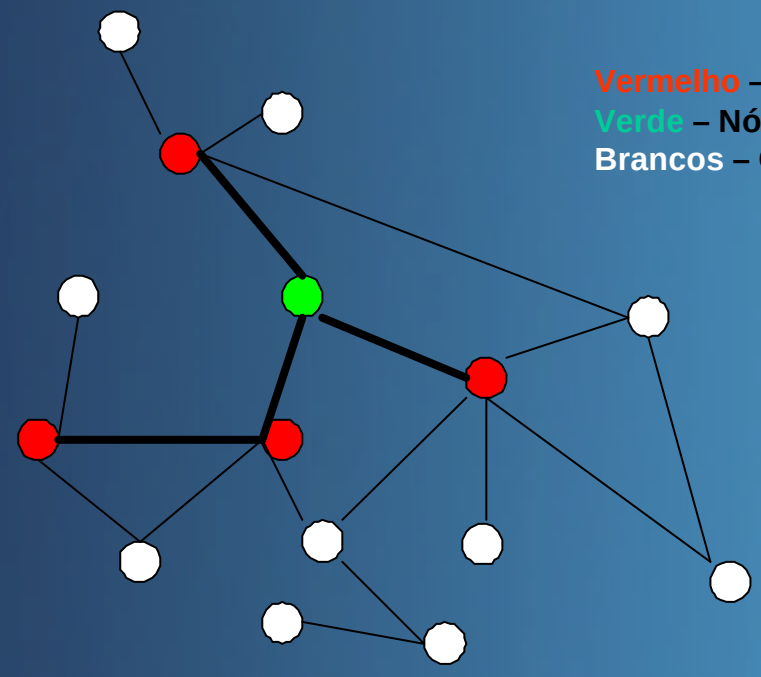


O Problema

Como descobrir a mais forte associação (e possíveis nós intermediários) entre dois ou mais nós ?

A Solução

Algoritmo de Caminho Mínimo identifica o caminho mínimo entre cada par de nós-fonte



Vermelho – Nós Fonte
Verde – Nós intermediários
Brancos – Outros Nós

(H. Chen, Coplink, 2004)

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

Construção das Redes Utilizadas para Pesquisa

Algoritmos propostos

Conceitos, Algoritmos, Indicadores, Avaliação

Recomendações dos autores

Apreciação

Tema da Tese

(Chen, Coplink, 2004)

Objetivos

Os autores propõem uma técnica para análise de redes que pode identificar os caminhos mais fortes nas associações entre duas ou mais entidades da rede criminal.

Avaliar a performance das técnicas empregadas em termos de:

- Efetividade – utilidade das associações
- Eficiência – velocidade do algoritmo

(*t* médio de execução - Chen & Xu, 2003)

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

Construção das Redes Utilizadas para Pesquisa

Algoritmos propostos

Conceitos, Algoritmos, Indicadores, Avaliação

Recomendações dos autores

Apreciação

Tema da Tese

(Chen, Coplink, 2004)

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

- **Revisão de Conceitos**

  Construção das Redes Utilizadas para Pesquisa

  Algoritmos propostos

  Conceitos, Algoritmos, Indicadores, Avaliação

Recomendações dos autores

Apreciação

Tema da Tese

Revisão de Conceitos - Link Analysis

Permite a descoberta de linhas de investigação "invisíveis"

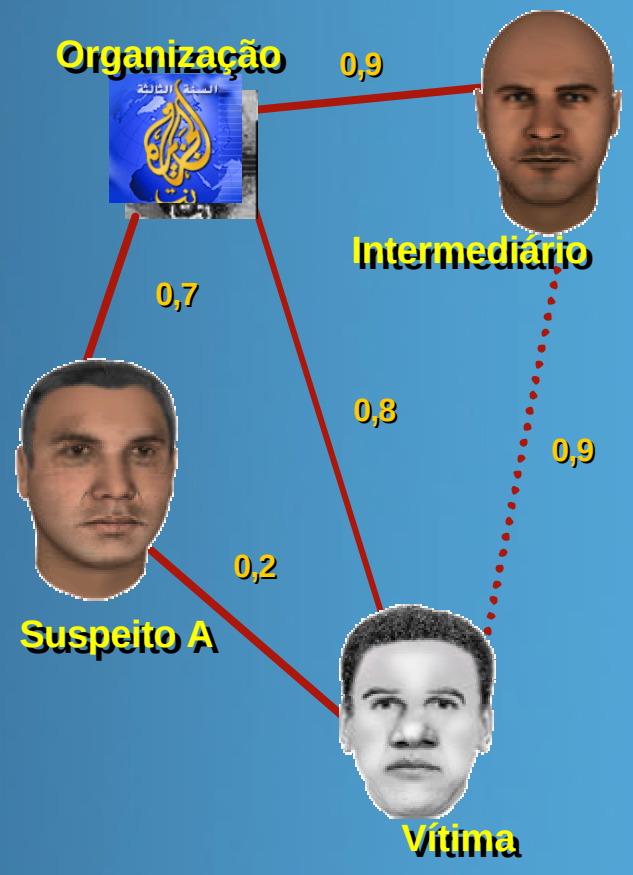
Ex:

Conexões podem evidenciar que determinado grupo de marginais pertencem a mesma cadeia de suprimento de drogas.

Em homicídios, a análise de conexões pode descobrir intermediários....

... conectando a vítima ao suspeito

...nega conhecer a vítima.



Revisão de Conceitos - Link Analysis - Entidades

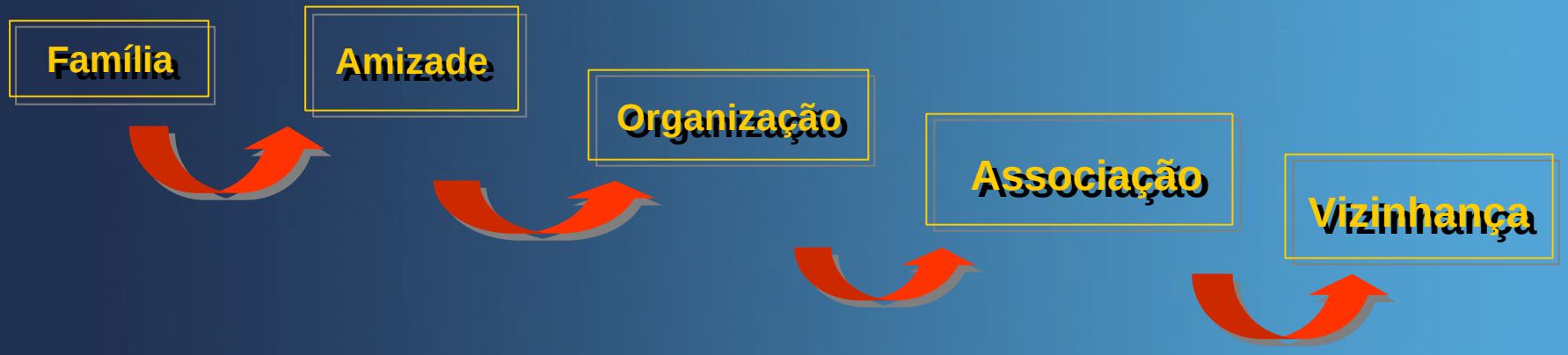
[Chen & Lynch, 1992]

Referências utilizadas para interelacionamento entre os elos de uma investigação

Revisão de Conceitos - Link Analysis - Entidades

Como estas entidades estão associadas?

Os criminosos encontram-se conectados:



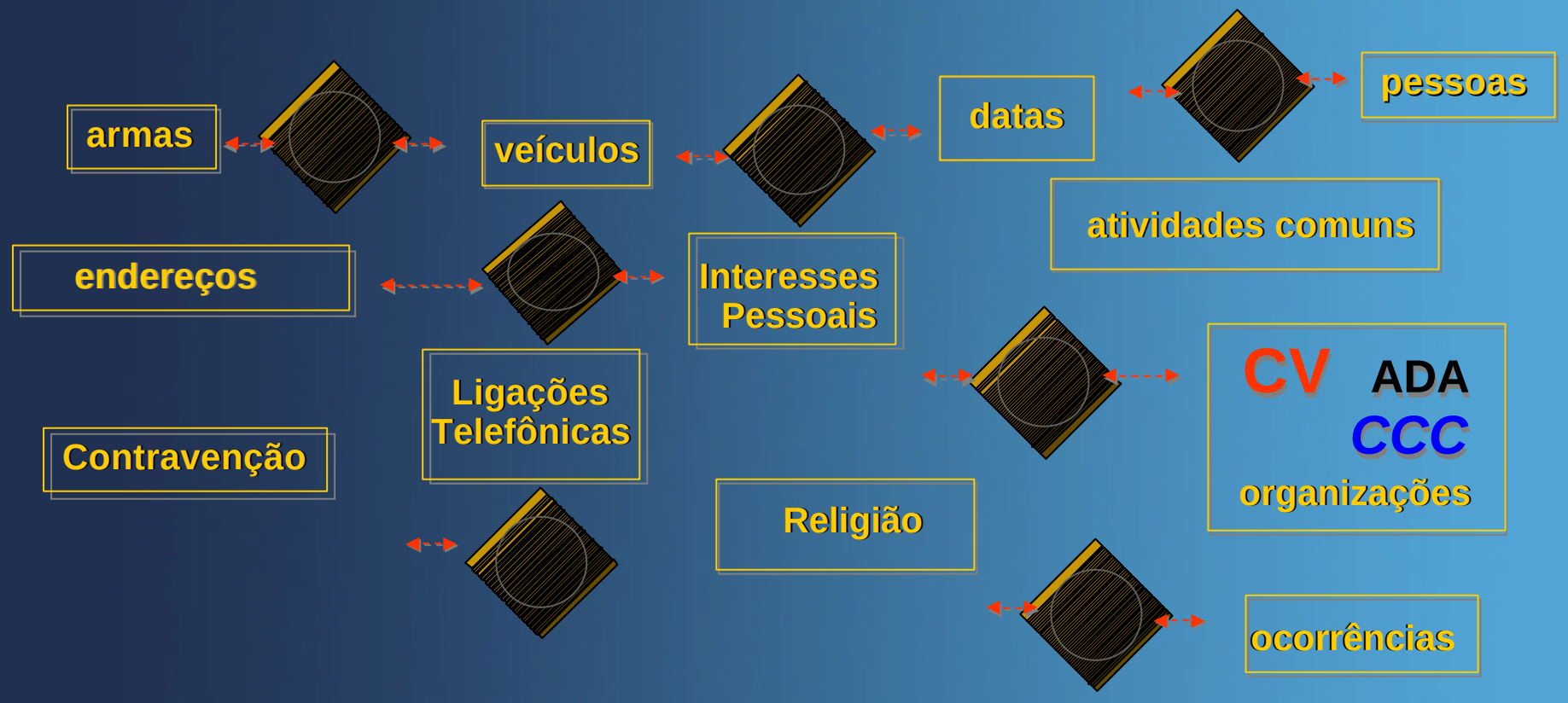
(Chen, Coplink, 2004)

Revisão de Conceitos - Link Analysis - Entidades

[*Chau , Xu, Chen, 2002]

Problemas gerados pela extração automática - links inúteis

Dualidades



Obstáculos à automação de conexões

- Integração dos dados extraídos
- Excesso de conexões
- Demandam tempo para avaliar a relevância de todos os Links e utilidades

(Chen, Coplink, 2004)

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

☞ **Construção das Redes Utilizadas para pesquisa**

Algoritmos propostos

Conceitos, Algoritmos, Indicadores, Avaliação

Recomendações dos autores

Apreciação

Tema da Tese

(Chen, Coplink, 2004)

Técnicas e ferramentas para extração automática de entidades e construção de redes criminais

- O maioria das ferramentas de análise de redes criminais apenas a representa visualmente, não executando análise [M Sparow, 1991]

Construção de Redes

- Bancos de Dados Estruturados : [Goldberg & Senator 1998]
- AZNP, Coplink Detect [Coplink, 2001]
- Documentos Textuais [Lee, 1998]

(H. Chen, Coplink, 2004)

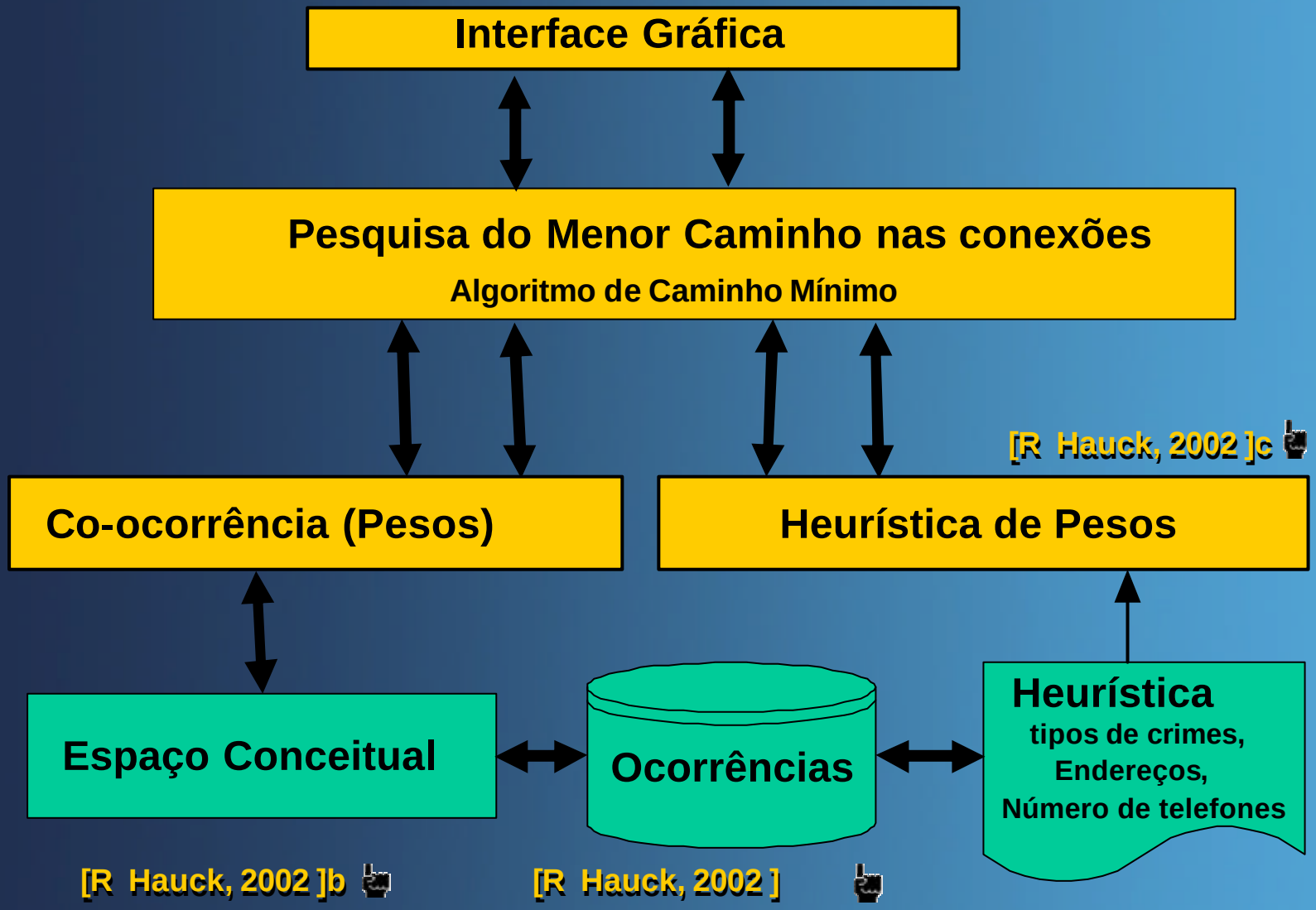
Resumo da construção das Redes utilizadas pelos autores para esta pesquisa

Modelagem:

Origem dos Documentos - Phoenix Pol Dept

- Seleção – Narcóticos & Sequestros (1 ano)
- Eliminados relatórios com 5 linhas ou menos
- Filtrados termos irrelevantes, dentre 3400 itens
- As frases/ nomes foram utilizadas como links
- As co-ocorrências foram calculadas
- Duas redes foram produzidas:

  Sequestros e Narcóticos



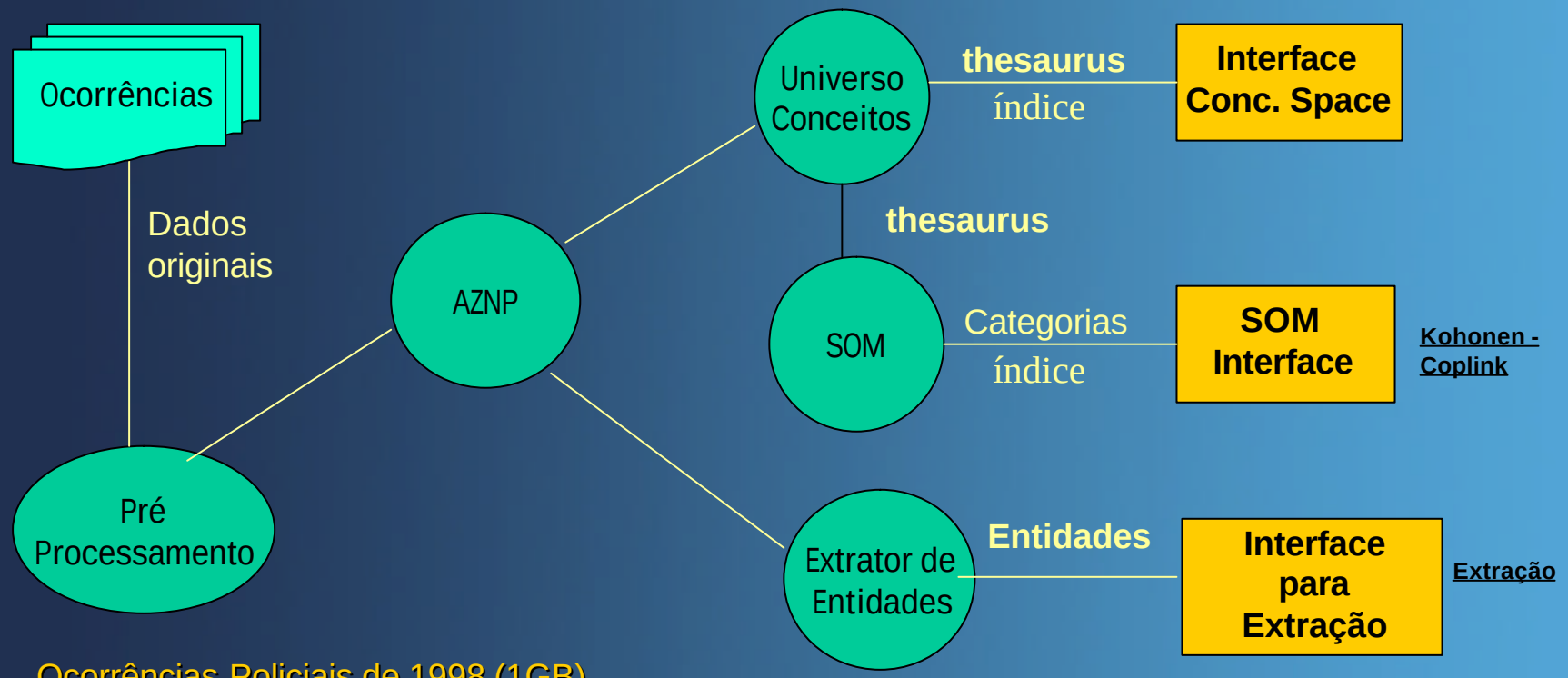
CrimeLink Explorer: Using Domain Knowledge to Facilitate Automated Crime Link Analysis
Lt. Jennifer Schroeder Tucson Police Department, Jie Xu University of Arizona, 2003

Construção de Redes

Exemplo de modelagem de Redes Automáticas

COPLINK Concept Space (textual)

- Usando AZNP para extração de Nomes e Frases
- Criando um Universo de Conceitos
- Transformando co-ocorrências em pesos



Ocorrências Policiais de 1998 (1GB)
Ocorrências com Sequestros (4.5MB)
- 271 ocorrências
- 280 conceitos (n) - 25,862 links (m)
- 92.4 links / conceito

Ocorrências com Narcóticos (38 MB)
- **3572 ocorrências**
- **4357 conceitos (n) - 733.752links (m)**
- **172.3 links / conceito**

(Chen, Coplink, 2004)

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

Construção das redes utilizadas para pesquisa

☞ **Algoritmos propostos**

Conceitos, Algoritmos, Avaliação

Recomendações dos autores

Apreciação

Tema da Tese

(Chen, Coplink , 2004)

Algoritimos de Caminho Mínimo entre um par de nós, citados pelos autores

- **Algoritmos tradicionais**
  - **Algoritmo Dijkstra** [Dijkstra, 1959] , [Cormen, 1999]
  - **Priority-First Search** [Helgason, 1993] , [Cormen, 1999]
  - **Floyd´s ou Dantzig´s** [Floyd, 1962] , [Dantzig, 1993]

- **Abordagem Neural**
  - **Rede Hopfield (Ali et al.,1993; Araujo, et al.,2000)** [Araújo et Al, 1994]

(Chen, Coplink , 2004)

- **Selecionados para cálculo do caminho mínimo entre dois nós fonte**

  PFS , Dijkstra/PFS bi - direcional [Helgason,1993]
  [Cormen, 1999] pp 149, 527

  Identificação do menor caminho de conexões dentre entidades na rede criminal.

- **Para simulação de operações típicas de conexões entre entidades na Rede Criminal;**

  BFS

  Permitir o benchmark com algoritmos de caminho mínimo

  [Cormen, 1999] pp 469

Como avaliar os algoritmos propostos pelos autores

Efetividade

- Avaliar a utilidade dos caminhos gerados por extratores

...Comparando os algoritmos PFS x BFS modificado
(para diferentes tipos de delitos).

Eficiência

- A atividade anti- criminal opera com prazos críticos.

  Algorítmos mais rápidos ajudam a redução do tempo de elucidação .

...analisando os algoritmos PFS x PFS modificado
(para diferentes tipos de delitos).

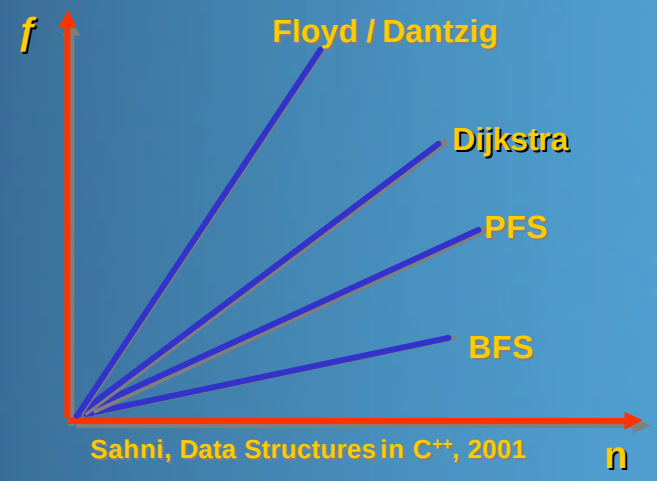
Complexidade dos algoritmos selecionados

	Dijkstra	PFS / PFS bi-direcional	BFS
Complexidade	$O(n^2)$	PFS $O(k(n+m)\log n)$, PFS bi-direcional $O(k^2(n+m)\log n)$	$O(n+m)$
	[Evans, 1992] pp36	[Evans, 1992] pp36	[Cormen, 1999]

Não Selecionados Complexidade $O(n^3)$

Floyd´s [Floyd,1962]

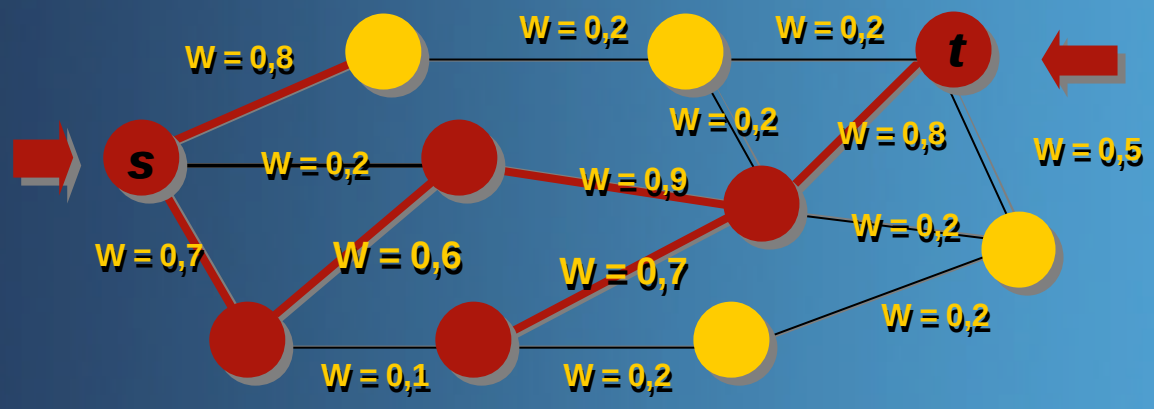
Dantzig´s [Dantzig, 1993]

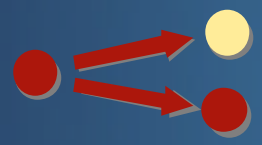


Sahni, Data Structures in C++, 2001

Modificações - O Algoritmo BFS

O algoritmo BFS pode localizar caminhos desde um nó fonte *s* até um nó *t*.



Modificação do BFS  nós sucessores.

Busca nós, cujo peso de conexão seja superior a determinado limite pré-estabelecido.

Garante os elos entre as mais fortes conexões.

Modificações - PFS

O Algoritmo é modificado para:

- calcular o caminho mínimo entre dois nós fonte, *s* - *t*

- calcular o caminho mínimo entre múltiplos (*k*) nós selecionados, (*k* - 1) repetições.

- permitir uma forte associação entre múltiplos nós (cada par selecionado estará fortemente associado)

O número total de caminhos mínimos localizados é *k* (*k* - 1) / 2.

[Cormen, 1999] pp 527

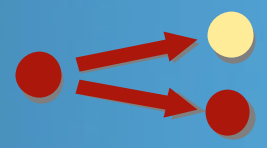
PFS bi-direcional (sem modificações)

Pesquisa ambos os terminais simultaneamente ➡ **6% nós**

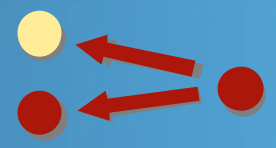
[Evans et Al apud Helgason, 1992] pp 92.

Um caminho mínimo com raiz em *s* e outro caminho mínimo com raiz em *t* progridem simultaneamente , em etapas alternadas.

a árvore enraizada em *s* cresce verificando os nós sucessores



a árvore enraizada em *t* cresce verificando os nós antecessores.

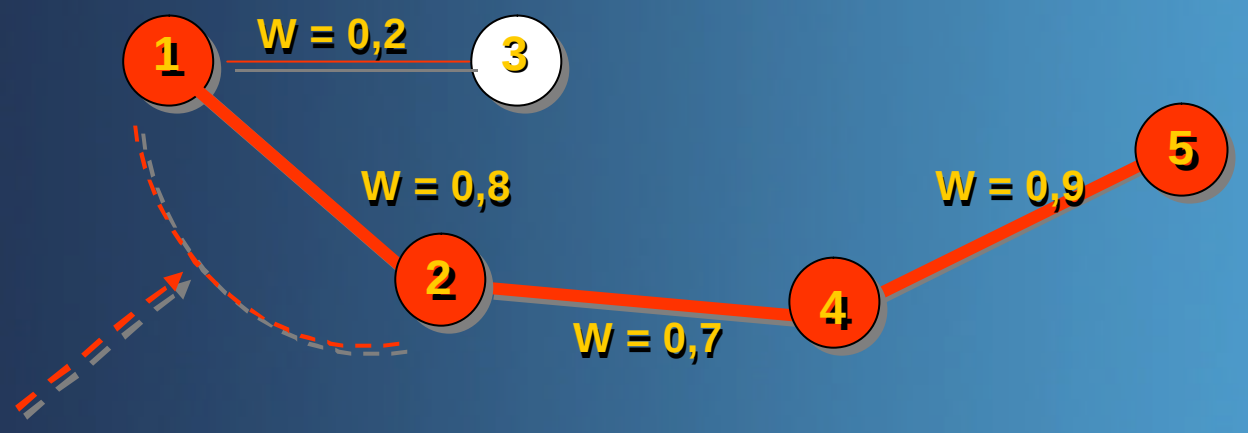


A pesquisa será encerrada quando ambas as árvores encontram um nó comum *r* , *tal que*

$d_r^s + d_r^t$ *seja mínima*

A Rede Transformada para Calculo de Conexões Fortes

As co-ocorrências (pesos) estão representadas por conexões fortes



O algoritmo está preparado para calcular o caminho mínimo na rede

... que corresponde à mais forte conexão

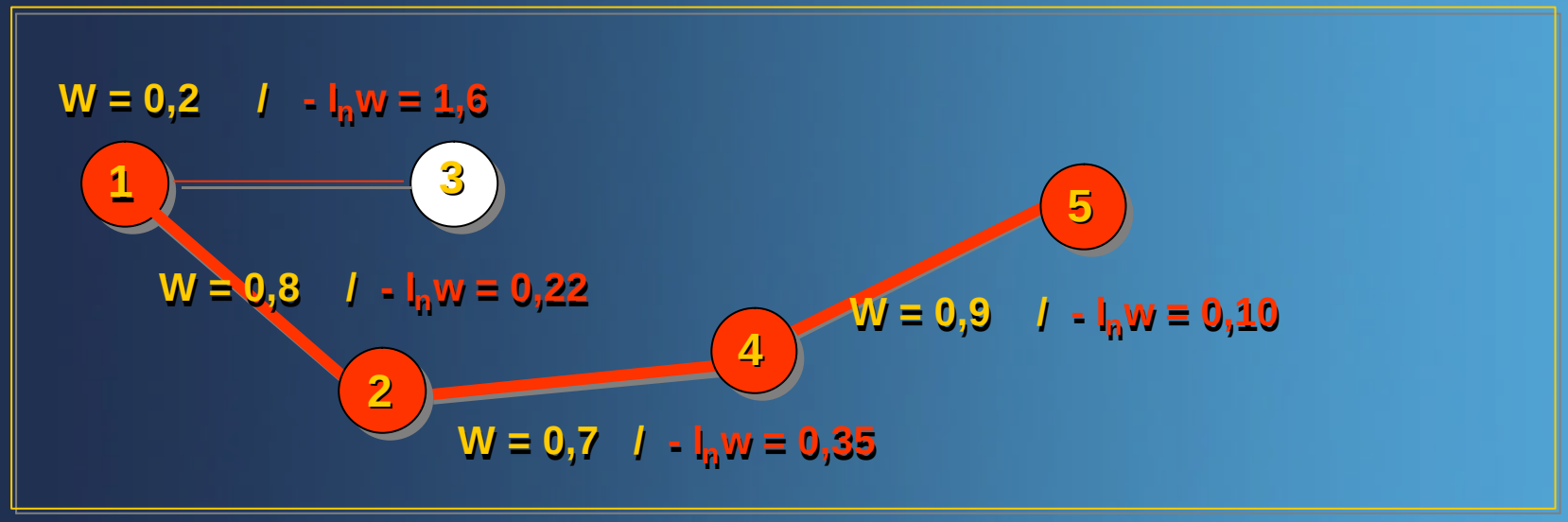
...uma transformação é executada na rede para compatibilizar estas premissas de cálculo

A Rede Transformada para Calculo de Conexões Fortes

A representação do peso $0 < w = 1$ na rede original é transformada, conforme abaixo:

l_{ij} o valor do link (i , j) ? A

$l = - \ln w$ $\qquad$ $- \ln w = 0$



O cálculo da conexão é dados pelo produtório dos pesos dos links intermediários entre u_k e u_l

Valor da Conexão entre u_k e u_l = ? $_k^l W_{k,l}$? $_1^5$ = 0,8 * 0,7 * 0,9 = 0,504

Avaliação dos Algoritmos

Estatísticas das redes pesquisadas pelos autores

Redes	ocorrências	Entidades extraídas	Tamanho rede (n)	Links (m)	Média de links p/ nó
Sequestro	271	95.328	280	25.862	92.4
Narcótico	3.572	861.516	4.257	733.572	172.3

Avaliação de Eficiência

- Simulação: k = 2, …, 5; conjuntos de nós selecionados (casos reais usam, em média, menos de 5 entidades)

- Gerados 100 casos randomicos, para cada rede, cada algoritmo

Média de execução em segundos para o cálculo do caminho mínimo

(os números entre parênteses representam o desvio padrão da amostra)

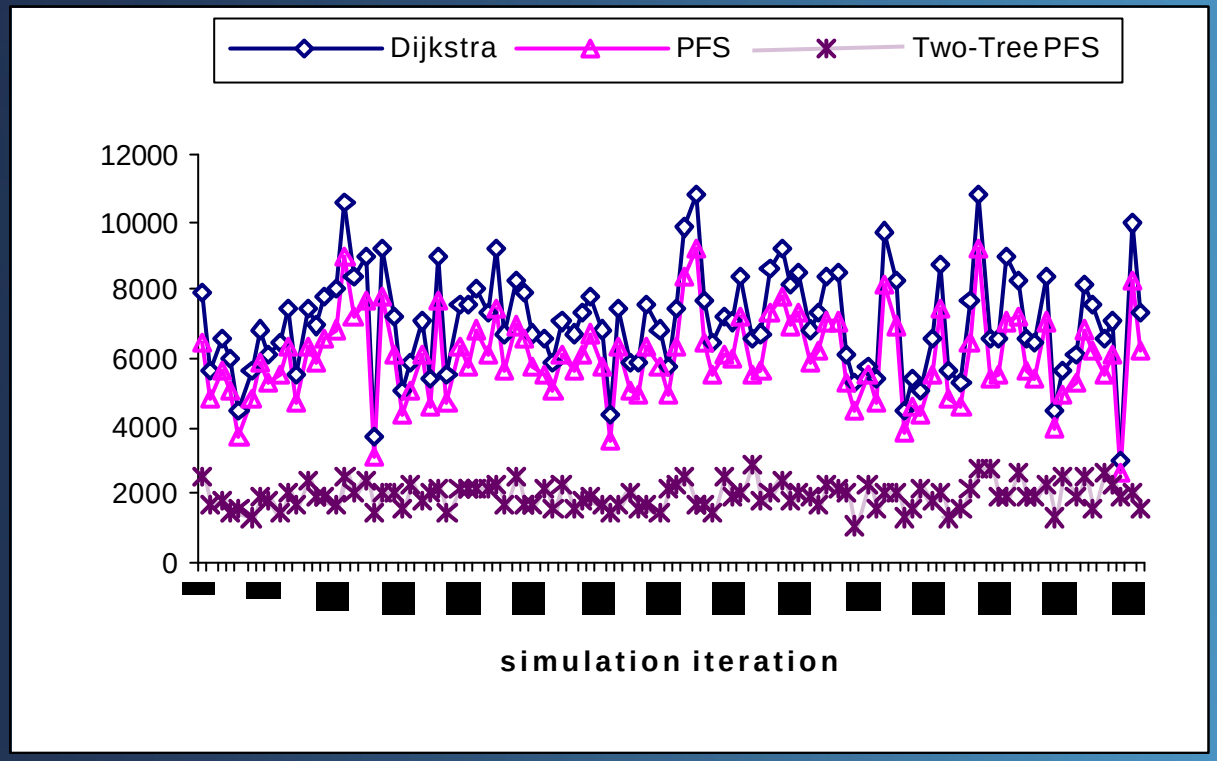
Algorithm	k = 2	k = 3	k = 4	k = 5
Dijkstra	1,170.4 (627.1)	3,388.2 (1162.7)	7,082.6 (1,517.3)	12,263.8 (1,517.3)
PFS	1,000.8 (543.1)	2,886.8 (967.8)	6,004.2 (1,259.7)	10,669.6 (2,092.8)
Two-tree PFS	351.3 (186.8)	946.2 (283.5)	1,940.0 (373.2)	3,447.9 (654.2)

Resultados para a Rede Sequestros

(H. Chen, Coplink, 2004)

Avaliação de Eficiência

Amostragem de tempos de processamento (k=5)

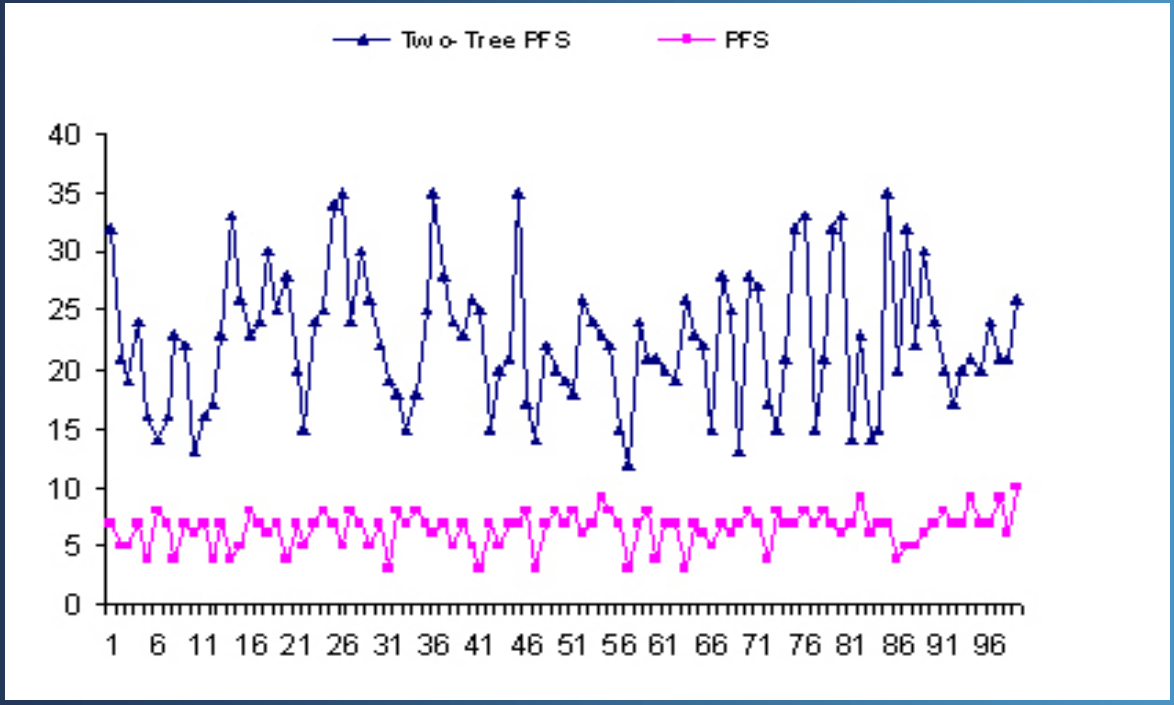


Resultados utilizando a Rede Sequestros

(H. Chen, Coplink, 2004)

Avaliação de Eficiência

Amostragem de tempos de processamento (k=5)



Resultados utilizando a Rede Narcotráfico

(H. Chen, Coplink, 2004)

Avaliação de Eficiência

Velocidade (1)

	PFS bi-direcional	PFS
Sequestros	+ RÁPIDO	
Narcoticos		+ RÀPIDO

Nós processados até geração do caminho mínimo

	PFS	PFS bi-direcional	Densidade Rede ñ direc
HELGASON (1993)	50%	6%	(2 m) número de Links / n (n -1) Links possíveis
Sequestros	52%	14,7%	0,66
Narcóticos	49,6%	3,9%	0,08

[Overlapping]

[S.Wasserman, 1994]

Avaliação de Eficiência

Velocidade (2)

PFS bi-direcional

PFS

INEFICIENTE

+ RÀPIDO

Pequenas e Densas

Grandes e Esparsas

Tratamento da fila temporária de etiquetagem

	Nós	PFS B.dir	PFS
Sequestros	280	120 %	47,9%
Narcóticos	4257	120 %	47,9%

Avaliação de Efetividade

Efetividade (utilidade)

Pequenas e Densas

Grandes e Esparsas

Redes	Média de Links nos caminhos		(*) Precisão	
	Sequestros	Narcoticos	Sequestros	Narcóticos
Caminho mínimo	1,40	2,06	66,7 %	71,4 %
BFS	1,73	12,50	30,0 %	16,7%

utilidade dos links

Especialistas analistas experientes da Polícia de Tucson

Utilidade dos links

$$(*)\ \text{Precisão} = \frac{\text{Número de caminhos úteis localizados por especialistas}}{\text{Número de caminhos úteis identificados por algoritmo}} \times 100\%$$

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

Construção das redes utilizadas para pesquisa

Algoritmos propostos

Conceitos, Algoritmos, Avaliação

Recomendações dos autores

Apreciação

Tema da Tese

(Chen, Coplink, 2004)

Recomendações dos autores

Indicadores de efetividade (caminhos úteis) podem ser melhorados, através de técnicas mais aprimoradas de extração de entidades

Entidades Nominais

~~Frases~~

AZNP – Extract Noum-Phrasing Tools - Tolle & Chen, 2000

Extração automática de entidades denominadas "Entidades Nominais" Nomes de pessoas, narcóticos, veículos e demais entidades relevantes

Extracting Meaningful Entities from Police Narrative Police – (Chau, Xu, Chen 2002)

Recomendações dos autores

Indicadores de efetividade (caminhos úteis) podem ser melhorados, através de técnicas mais aprimoradas de extração de entidades

Melhoria na Heurística de Extração

[Lee, 1998]

Máquina de aprendizado - arquitetura Back Propagation, obtém significado mais provável para a entidade (regras léxicas obtidas na fase de treinamento), de acordo com a frase analisada.

Extracting Meaningful Entities from Police Narrative Police – (Chau, Xu, Chen 2002)

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

Construção das redes utilizadas para pesquisa

Algoritmos propostos

Conceitos, Algoritmos, Avaliação

Recomendação dos Autores

Apreciação

Tema da Tese

(Chen, Coplink, 2004)

Apreciação (1)

Problemas citados pelos autores como falhas nos indicadores de utilidade:

a. Ambiguidade ;

b. Significado relativo - tipicação do crime

c. Baixa utilidade entre entidades, mesmo se presentes no mesmo documento

Extração → Crimes primários → Baixa co-ocorrência

Muitos homicídios, estupros e roubos são praticados por réus primários ou primários ainda não capturados, sem conexões registradas em boletins de ocorrências

Apreciação (2)

A base de "Irrelevâncias" do filtro inicial pode diferir de crime para crime

[Lee, 1998]

Apreciação (3)

O “Modus Operandi” (*) , também conhecido como“Dinâmica” constitui-se como entidade relevante na formação de conexões e caminhos úteis.

(*) Modus operandi é uma expressão em latim (algumas vezes também representada por sua abreviatura: MO), cujo significado, no trabalho policial, refere-se à descrição do padrão de atuação de um determinado criminoso ou grupo deles (F.Dantas, F. Moreira, 2006*)*

Proposição dos Autores

Resumo da proposição

Objetivos da Proposta

Desenvolvimento da Pesquisa

Revisão de Conceitos

Construção das redes utilizadas para pesquisa

Algoritmos propostos

Conceitos, Algoritmos, Avaliação

Recomendação dos Autores

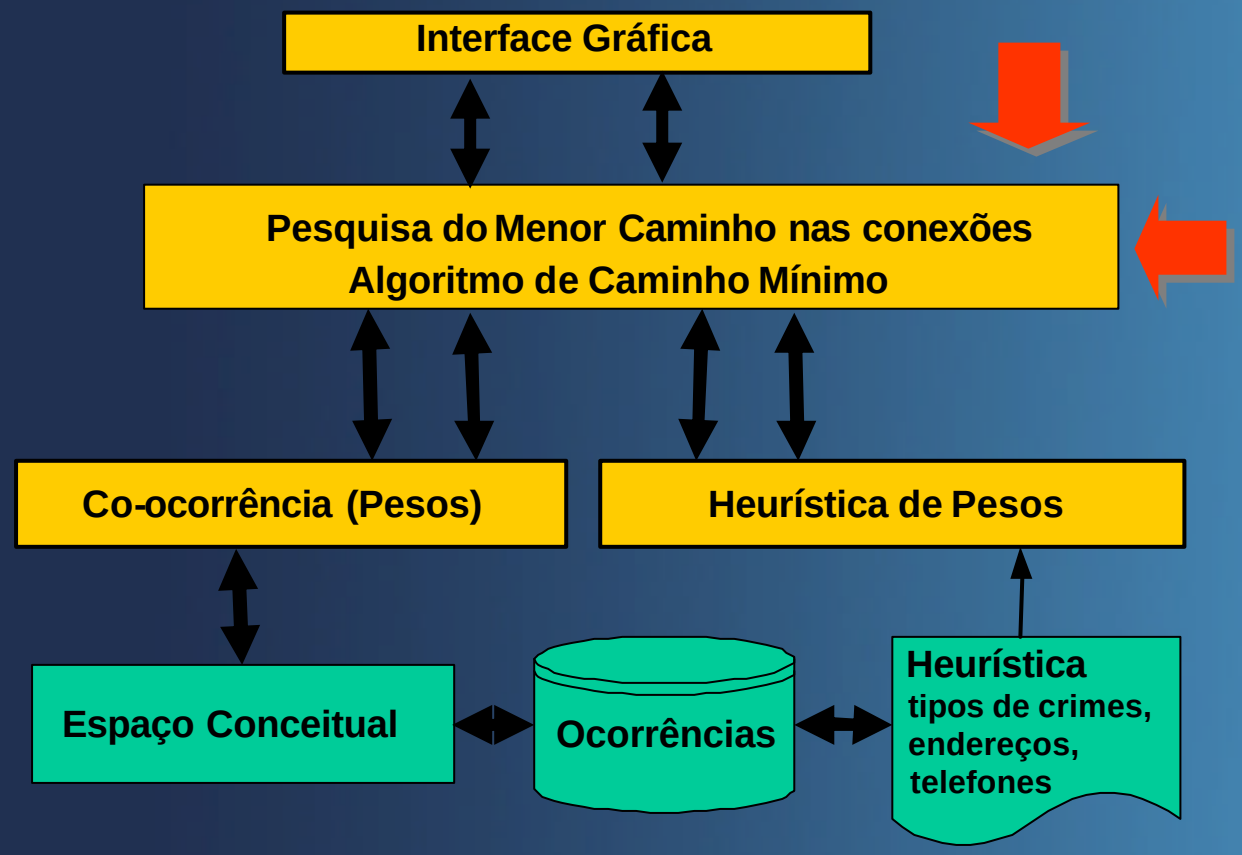
Apreciação

Tema da Tese

(Chen, Coplink, 2004)

Tema da Tese

As base de ocorrências brasileiras são geralmente volumosas, textualmente longas e desestruturadas



Pesquisar algoritmo adequado:

efetivo - extração de utilidades em bases extensas

eficaz para modelos hiper volumosos

Bibliografia

D. McAndrew, The structural analysis of criminal networks, in: D. Canter, L. Alison (Eds.), The Social Psychology of Crime: Groups, Teams, and Networks, Offender Profiling Series, Aldershot, Dartmouth, vol. III, 1999.

E. Dijkstra, A note on two problems in connection with graphs, Numerische Mathematik 1 (1959) 269– 271.

F. Araujo, B. Ribeiro, L. Rodrigues, A neural network for shortest path computation, IEEE Transactions on Neural Networks 12 (5) (2001) 1067–1073.

G. Dantzig, On the shortest route through a network, Management Science 6 (1960) 187– 190.

H. Chen, K.J. Lynch, Automatic construction of networks of concepts characterizing document databases, IEEE Transactions on Systems, Man and Cybernetics 22 (5) (1992) 885– 902.

H.G. Goldberg, R.W.H. Wong, Restructuring transactional data for link analysis in the FinCen AI System, Proceedings of 1998 AAAI Fall Symposium on Artificial Intelligence and Link Analysis, AAAI Press, Menlo Park, CA, 1998.

H.G. Goldberg, T.E. Senator, Restructuring databases for knowledge discovery by consolidation and link formation, Proceedings of 1998 AAAI Fall Symposium on Artificial Intelligence and Link Analysis, AAAI Press, Menlo Park, CA, 1998.

J. Evans, E. Minieka, Optimization Algorithms for Networks and Graphs, Marcel Dekker, New York, 1992.

K.M. Tolle, H. Chen, Comparing noun phrasing techniques for use with Medical Digital Library tools, J. A. Society for Information Science 51 (2000).

M. Ali, F. Kamoun, Neural networks for shortest path computation and routing in computer networks, IEEE Transactions on Neural Networks 4 (5) (1993)

M. Chau, J. Xu, H. Chen, Extracting meaningful entities from police narrative reports, Proceedings of the National Conference on Digital Government Research (Los Angeles, CA), 2002, pp. 271– 275.

M.K. Sparrow, The application of network analysis to criminal intelligence: an assessment of the prospects, Social Networks 13 (1991) 251– 274.

P. Klerks, The network paradigm applied to criminal organizations: theoretical nitpicking or a relevant doctrine for investigators? Recent developments in The Netherlands, Connections 24 (3) (2001) 53–65.

R. Lee, Automatic information extraction from documents: a tool for intelligence and law enforcement analysts, Proceedings of 1998 AAAI Fall Symposium on Artificial Intelligence and Link Analysis, AAAI Press, Menlo Park, CA, 1998.

R.V. Hauck, H. Atabakhsh, P. Ongvasith, H. Gupta, H. Chen, Using coplink to analyze criminal-justice data, IEEE Computer 35 (3) (2002) 30– 37.

R.V. Helgason, J.L. Kennington, B.D. Stewart, The one-toone shortest-path problem: an empirical analysis with the two-tree dijkstra algorithm, Computational Optimization and Applications 1 (1993) 47– 75.

R.W. Floyd, Algorithm 97: shortest path, Communications of the ACM 5 (6) (1962) 345– 370.

S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Cambridge Univ. Press, Cambridge, 1994.

T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms, MIT Press, Cambridge, MA, 1991.

V.E. Krebs, Mapping networks of terrorist cells, Connections 24 (3) (2001) 43–52.

W.R. Harper, D.H. Harris, The application of link analysis to police intelligence, Human Factors 17 (2) (1975) 157– 164.

Bibliografia – outras referências

CrimeLink Explorer: Using Domain Knowledge to Facilitate Automated Crime Link Analysis
Lt. Jennifer Schroeder Tucson Police Department, Jie Xu University of Arizona, 2003

H. Chen, W. Wahg. Coplink, Disponível em http://ai.bpa.arizona.edu/hchen/docs/NIJ-DM-DC2002.ppt , 2003

F.Dantas, F. Moreira. Vínculos Criminais. Disponível em http://www.peritocriminal.com/artigos/investcrim.htm , 2007

I. Martins. VisualGrafus. PUC-Rio, 2005/2006

Obrigado

Isnard Martins