



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Mineração de Links

Iuliana G. S. Rodrigues

iulianas@gmail.com

Pedro Shiguihara-Juárez

p.shiguihara@gmail.com

Jorge Valverde-Rebaza

jorge.carlos14@gmail.com

2 de junho de 2011



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Mineração de Links

- Definição: Técnica de mineração de dados que consideram explicitamente esses links quando constroem modelos preditivos ou descritivos de dados.
- *Links* ou relacionamentos são únicos. Objetivos similares a Data Mining.
- Exemplo: Relações entre as entidades: Sintoma - Doença, Links em páginas da WEB, Amizades, Interesses similares, etc.
- Explora as informações representadas pelas Relações (ou Links)



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Mineração de Links

- **Data Mining:**
 - Busca extrair conhecimento útil dos dados
 - Eficiente ao analisar entidades independentes
 - Comete muitos erros quando trata entidades interdependentes como sendo independentes
- **Link Mining:**
 - Objetivos similares a Data Mining
 - Eficiente ao analisar entidades interdependentes
 - Explora as informações representadas pelas Relações (ou Links)



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Mineração de Links

- Link Mining é a intersecção de:
- Análise de links
- Mineração da Web e de Hipertextos
- Aprendizagem relacional
- Programação em Lógica Indutiva
- Mineração de grafos



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Ranking: Definição

- Definição: É uma tarefa do *link mining* utilizada para explorar a estrutura de um grafo para ordenar ou priorizar um conjunto de objetos dentro de um grafo. (Getoor et. al)
- *PageRanking* e *HITS*: Algoritmos mais notáveis nessa abordagem.
- O objetivo é explicar a estrutura de links de um grafo para ordenar ou priorizar o conjunto de objetos dentro do grafo. (Foca em grafos com um objeto simples e tipos de links simples.)
- A definição da centralidade depende do que é considerado importante para a aplicação



Introdução

Tarefas com
Objetos -
Vértices

**Links Baseados
em Ranking de
Objetos**

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Ranking: Aplicações

- (Page e Brin, 1998) Algoritmo para ranking de páginas web

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteccão de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

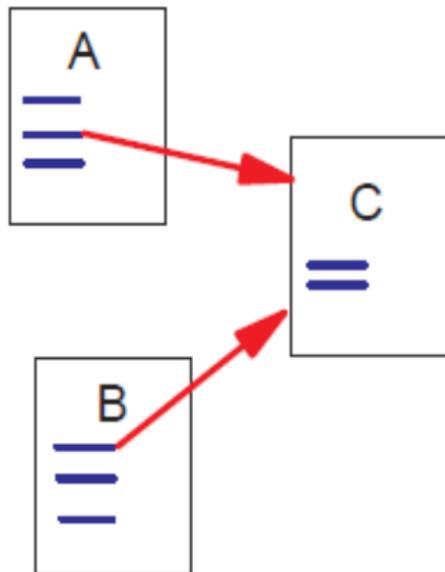


Figure 1: A and B are Backlinks of C

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Ranking: Exemplos

- Um grande número de algoritmo tem sido proposto para uma variedade desses temas:
- Chakrabarti et al 2005 - propõe modificações de hits que exploram a página web conteúdo para de peso e links com base na relevância.
- Haveliwala e Jeh e Widom, 2003 - propõem um algoritmo sensível baseado em PageRank que identifica páginas autoritativa no momento da consulta.
- Cohn e Cahng,2003 - introduz uma analogia probabilística de hits, baseado na indexação semântica



Classificação de links

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Classificação de links

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Classificação de links

- 1** Objetivo: classificar os membros de um conjunto composto de objetos conectados de um conjunto finito de valores categóricos.
- 2** Identificação com base em relação previamente observada.
- 3** Dois vértices relacionados tendem a apresentar correlação na classificação
- 4** Objetos estruturalmente idênticos que são predefinidos para ser IID (Independent, Identically Distributed)

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Classificação: Aplicações

- Identificar indivíduos como parte de uma rede terrorista com base em relações observadas
- Relações específicas com outros membros da organização podem ser indícios de participação



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Classificação de links: Aplicações

- 1 Taskar et al, 2002 estende a abordagem de Laferty et al.'s para o caso onde os dados no grafo são grafos arbitrários.
- 2 Neville and Jensen, 2000 propuseram algoritmo simples para classificar dados corporativos com esquemas ricos que produziam grafos com objetos heterogêneos, cada um com seu próprio conjunto de características distintas.



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Detecção de grupo

- 1 Definição
- 2 Detecção de grupo em Redes Sociais

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Conclusões

Referências

Definição

- 1 O objetivo da detecção do grupo é fazer clusterização dos nós no grafo em grupos que compartilham características comuns (Getoor et. al)
- 2 **O objetivo é extrair conhecimento valioso que podem apresentar padrões ocultos (Adibi et. al)**

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Definição

- 1 O objetivo da detecção do grupo é fazer clusterização dos nós no grafo em grupos que compartilham características comuns (Getoor et. al)
- 2 O objetivo é extrair conhecimento valioso que podem apresentar padrões ocultos (Adibi et. al)
- 3 **Relação de instâncias de algum tipo de entidade sobre outras semelhantes (Ozgul et. al)**



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Definição

- 1 O objetivo da detecção do grupo é fazer clusterização dos nós no grafo em grupos que compartilham características comuns (Getoor et. al)
- 2 O objetivo é extrair conhecimento valioso que podem apresentar padrões ocultos (Adibi et. al)
- 3 Relação de instâncias de algum tipo de entidade sobre outras semelhantes (Ozgul et. al)
- 4 **Tenta detectar subgrupos coesos sobre os quais há uma relação forte, direta, intensa e frequente (Ozgul et. al)(Wasserman et al, (Wasserman and Faust))**

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

Fases

Segundo Wang et. al, a tarefa de detecção de grupo tem duas fases:

- 1** Criar grupos por meio de uma semente
- 2** Expandir os grupo existente por teste de potenciais membros



Detecção de grupo: Definição

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Aplicações

- 1 Mineração de dados
- 2 Análise de Redes Sociais
- 3 Teoria de grafos



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

**Detecção de
grupo**

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Análise de Redes Sociais: GAP

- 1 Um grande problema na área de Link Mining é a descoberta de estruturas organizacionais ocultas (Adibi J. et. al)(Ozgul et. al)**

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

**Detecção de
grupo**

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Análise de Redes Sociais: GAP

- 1 Um grande problema na área de *Link Mining* é a descoberta de estruturas organizacionais ocultas (Adibi J. et. al)(Ozgul et. al)
- 2 **Descoberta de membros de grupo NÃO conhecidos e identificação completa de grupos NÃO conhecidos (Adibi J. et. al)**

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

**Detecção de
grupo**

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Análise de Redes Sociais: Evidências

- Comunicação de membros de grupo
- Transações comerciais
- Relações familiares, etc.

Métodos escaláveis: Um caso

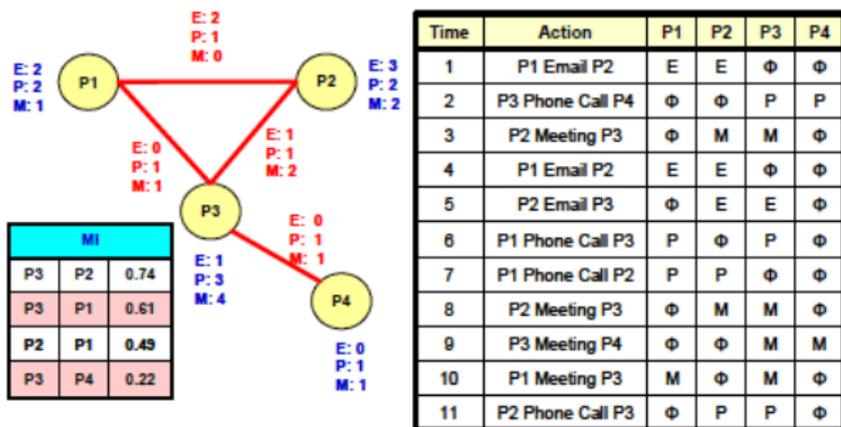


Figura: Obtenção de MI entre duas variáveis. Onde E:email, M:messages, P:phone.

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

**Detecção de
grupo**

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Caso: Bursa (Detecção)

- 1 Um modelo de detecção de grupos criminosos (*Offender Group Detection Model-OGDM*) foi testado em Bursa (Turquia) (Ozgul et. al)
- 2 Os resultados do OGDM foram que 34 pessoas estavam no grupo criminal de Bursa
- 3 Sete deles sem antecedentes criminais

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

**Detecção de
grupo**

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Caso: Bursa (Depois)

- 1** Depois de uma operação policial, a polícia prendeu aos criminosos
- 2** A rede real tinha 20 pessoas, das quais 3 não tinham antecedentes
- 3** A polícia recuperou U.S. \$ 200,000 em bens roubados e U.S. \$ 180,000 em dinheiro

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

**Detecção de
grupo**

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Métodos escaláveis: O desafio central

- O desafio é desenvolver métodos escaláveis que possam explorar incrementalmente grafos complexos para a obtenção e descoberta de conhecimento (Adibi J. et. al)(Getoor et. al)(Ozgul et. al)

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

**Detecção de
grupo**

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Métodos escaláveis: Um caso

- Aidbi et al.(Adibi J. et. al) propuseram um localizador de grupo (KOJAK Group Finder)
- Um posicionamento dos grupos era fixada
- Uma expansão desses grupos utilizando técnicas baseadas em conhecimento para acrescentar mais candidatos
- Depende das interações que mostram possíveis associações



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteccção de
grupo

**Resolução de
entidade**

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Resolução de entidade: Definição

- É a identificação do conjunto de objetos em um domínio



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Resolução de entidade: Definição

- Uma entidade pode ser qualquer objeto do mundo real: pessoas, lugares, coisas entre outras
- A resolução de entidade é o processo de determinar se dois referências do mundo real estão se referendo ao mesmo o diferentes objetos
- No contexto das bases de dados, a resolução de entidade é o processo de identificar quais registros numa base de dados refere-se à mesma entidade do mundo real, o que envolve medir a similaridade entre cada par de registros

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

**Resolução de
entidade**

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Resolução de entidade: Definição

- $E = \{e_1, e_2, \dots, e_m\}$ (Entidades)
- $R = \{r_1, r_2, \dots, r_n\}$ (Referências)
- O objetivo é agrupar corretamente as representações em R co-relacionadas, isto é, que se refere à mesma entidade



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

**Resolução de
entidade**

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Resolução de entidade: Definição

Exemplos deste problema são encontrados em:

- Bases de dados (integração de dados, duplicação)
- Problema relacionados com duplicação de dados



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

**Resolução de
entidade**

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Resolução de entidade: Abordagens

O abordagem poder ser visto segundo a quantidade de registros que são comparados

- Resolução independente
- Resolução por Naive relacional
- Resolução coletiva

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Resolução de entidade: Abordagens

O abordagem poder ser visto segundo a quantidade de registros que são comparados

- Resolução independente: a similaridade $sim_A(r_i, r_j)$ é calculado para cada par de referências r_i e r_j baseados em seus atributos

Resolução de entidade: Abordagens

Uma abordagem pode ser vista segundo a quantidade de registros que são comparados

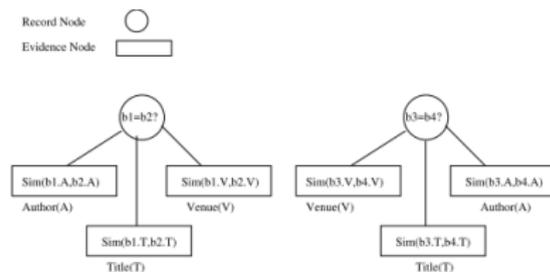
- Resolução independente
- Resolução por Naive relacional: a forma mais simples para usar relações para a resolução de entidades é tratar referências relacionadas como atributos adicionais para a associação. (Ex: considerar os nomes de seus coautores para um autor)

Resolução de entidade: Abordagens

O abordagem poder ser visto segundo a quantidade de registros que são comparados

- Resolução independente
- Resolução por Naive relacional
- Resolução coletiva: o objetivo é fazer uma *clusterização* das referências, de modo que só as referências que pertencem à mesma entidade sejam atribuídas ao mesmo *cluster*

Resolução de entidade: Abordagens



(a) Each pairwise decision considered independently

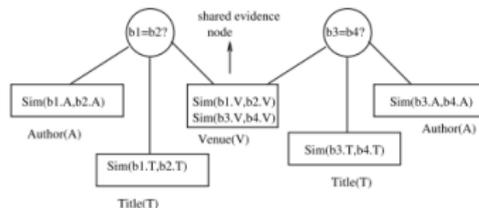


Figura: Forma em que trabalha o modelo independente e o modelo

Resolução de entidade: Abordagens

		Paper		Venue	
		indep	joint	indep	joint
Citeseer	constraint	88.9	91.0	79.4	94.1
	reinforce	92.2	92.2	56.5	60.1
	face	88.2	93.7	80.9	82.8
	reason	97.4	97.0	75.6	79.5
	Micro Avg.	91.7	93.4	73.1	79.1
Cora	kibl	92.9	93.3	93.6	99.3
	fahl	95.5	95.0	87.3	99.7
	utgo	79.9	84.0	51.7	60.4
	Micro Avg.	89.4	90.8	77.5	84.5

Figura: Tabela que mostra os experimentos de Culotta et. al, utilizado a resolução de entidades por *clusters* e de forma independente



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Predição de links: Definição

- Permite a estimação da existência de um link entre dois nós, baseado nos links e os atributos dos nós (Lü and Zhou).



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Predição de links: Classificação

- 1** O link existe mas é desconhecido para a amostragem das redes (Lü and Zhou).
- 2** O link pode existir no futuro da evolução das redes



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

**Predição de
links**

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Predição de links: Classificação

- 1** O link existe mas é desconhecido para a amostragem das redes (Lü and Zhou).
 - Redes de interações de proteína-proteína
 - Redes metabólicas
- 2** O link pode existir no futuro da evolução das redes
 - Redes sociais

Predição de links: Exemplo

- 1 Seja uma rede social: $G = (V, E)$. Cada aresta $e = (u, v) \in E$ é uma interação entre u e v num tempo $t(e)$ (Liben-Nowell and Kleinberg)
- 2 Múltiplas interações são feitas entre u e v como arestas paralelas em diferentes tempos.
- 3 Cada link tem um peso: $score(x, y)$
- 4 Um *Ranking* é obtido em função dos $score(x, y)$ para cada par de nós (Medida de similaridade o proximidade entre os nós x e y)



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

**Predição de
links**

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Predição de links: Técnicas

As técnicas podem ser (Song et. al.):

- 1 Preditores de links básica
- 2 Preditores de links composta

Predição de links: Técnicas

1 Preditores de link básica

- Consiste de uma medida de proximidade e um limiar T
- Um par de nós $(x, y) \notin E$ serão parte do conjunto de arestas se a proximidade entre eles é grande.
- $[x, y] \geq T$

2 Preditores de link composta

- Utiliza técnicas de Aprendizado de Máquina para fazer predições com várias medidas de proximidade

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Predição de links: Medidas de proximidade

1 Score: *Vizinhos comuns*

- $CN[x, y] = |\Gamma(x) \cap \Gamma(y)|$
- Há mais probabilidade de que x e y sejam amigos, se tem mais amigos em comum

Predição de links: Medidas de proximidade

1 Score: *Jaccard's coefficient and Adamic/Adar*

- $\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(\Gamma(z))}$
- Tenta medir o tamanho da interseção de dois vizinhos
- Poder dar mais peso ao nodo comum com menor número de amigos
- Mede o número de vizinhos de x e y comparado ao número de nós que são ou vizinhos com x ou vizino com y :

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Predição de links: Básica

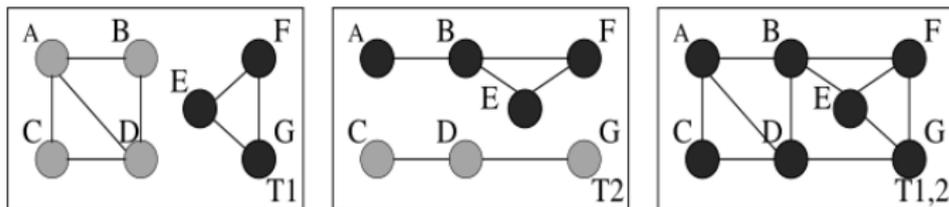


Figura: Evolução de um grafo G , onde os dois primeiros grafos são *snapshots* em tempo t_1 e t_2 respectivamente, e a seguinte *snapshot* é uma junção dos dois grafos anteriores

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Predição de links: Composta

$l = 1$		$l = 2$		$l = 3$	
x_i	w_i	$x_i \rightarrow x_j$	w_{ij}	$x_i \rightarrow x_j$	w_{ij}
a	4	$a \rightarrow b$	2	$a \rightarrow b \rightarrow c$	1
b	5	$a \rightarrow c$	1	$a \rightarrow b \rightarrow e$	1
c	4	$b \rightarrow c$	3	$a \rightarrow c \rightarrow d$	1
d	3	$b \rightarrow e$	1	$b \rightarrow c \rightarrow b$	1
e	1	$b \rightarrow g$	1	$b \rightarrow c \rightarrow d$	1
f	2	$c \rightarrow b$	1	$b \rightarrow c \rightarrow f$	1
g	1	$c \rightarrow d$	1	$b \rightarrow e \rightarrow d$	1
		$c \rightarrow f$	1	$c \rightarrow b \rightarrow g$	1
		$d \rightarrow f$	1	$c \rightarrow d \rightarrow f$	1
		$e \rightarrow d$	1	$c \rightarrow f \rightarrow a$	1
		$f \rightarrow a$	1		

Figura: Modelo de Markov para a navegação dos usuários na web

Predição de links: Composta

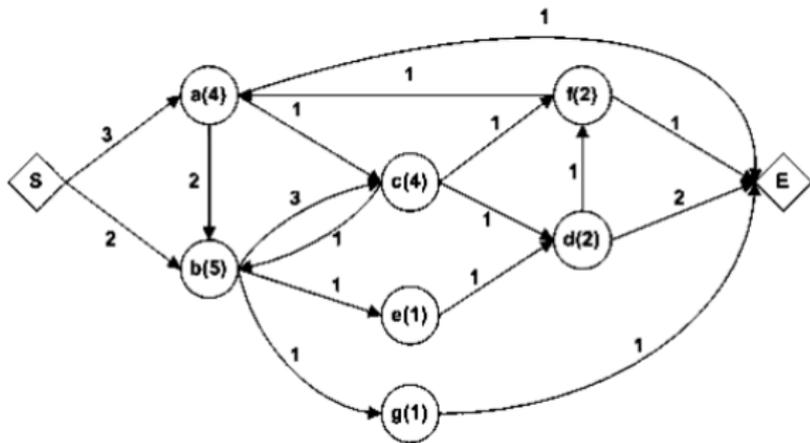


Figura: Onde os números em parênteses nos nós são a quantidade de visitas à site web, além, os pesos das arestas são o número de vezes que o link foi visitado



Tarefas relacionadas com Grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

**Tarefas com
Grafos**

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Tarefas relacionadas com Grafos



Tarefas relacionadas com Grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- Tarefas executadas em grafos a partir das informações dos links.
- Grafos são as estruturas de dados de caracter universal que permitem o modelagem de diferentes entidades e suas relações.
- Os algoritmos de mineração de grafos permitem o incremento ou entendimento das informações representadas pelos conjuntos de dados de grafos, as quais podem formar estruturas muito complexas.



Descoberta de Sub-grafos

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- A área de mineração de dados que tem relação com a mineração de links é o trabalho de descoberta de sub-grafos.
- É a busca de ocorrências de sub-grafos comuns ou interessantes em um conjunto de grafos.
- A maioria das propostas foco no uso das características topológicas de redes estáticas.
- Num modelo do mundo real leva em conta componentes dinâmicos e as interações entre objetos que ocorrem a cada certo período de tempo. Eles são nomeados grafos dinâmicos.

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

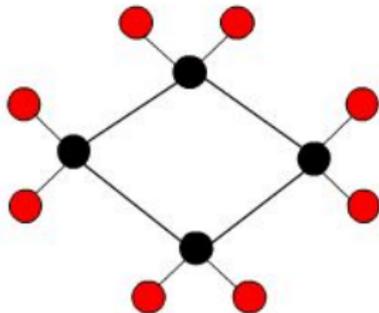
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

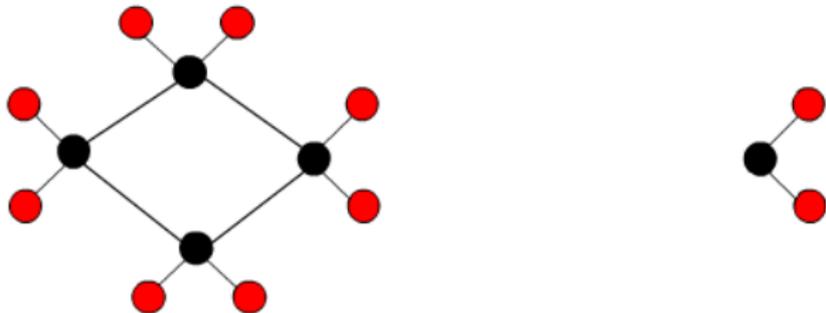
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

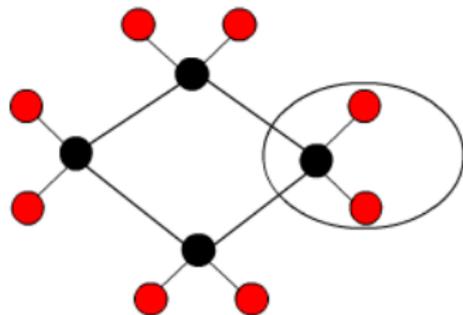
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteccão de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

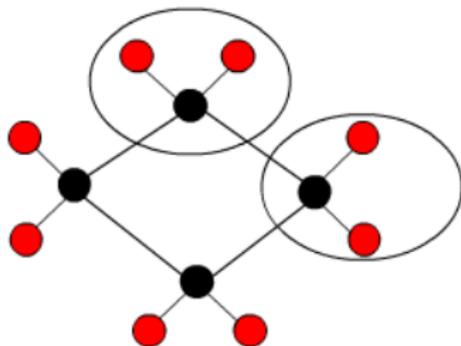
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

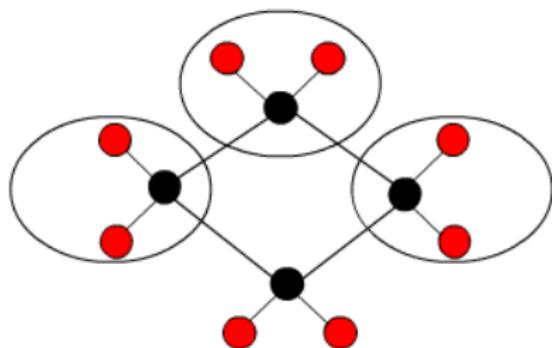
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

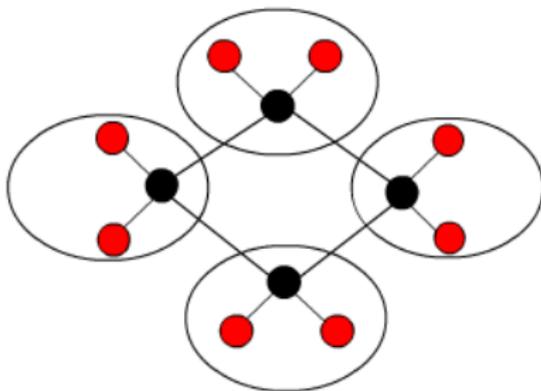
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- Os grafos dinâmicos ocorrem em muitas aplicações do mundo real:
 - Em Biologia é comum o modelagem das interações das proteínas como redes, onde cada vértice é uma proteína e dois vértices são conectados por uma aresta se as proteínas correspondentes têm relação entre elas.
 - Em Química é comum o modelagem das interações dos átomos como redes, onde cada vértice é um átomo e as arestas correspondentes são as relações entre os átomos (Deshaspe, et. al).
 - Em redes sociais o modelagem como redes é com pessoas como vértices, e as arestas são os contatos entre as pessoas em momentos específicos de tempo.

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

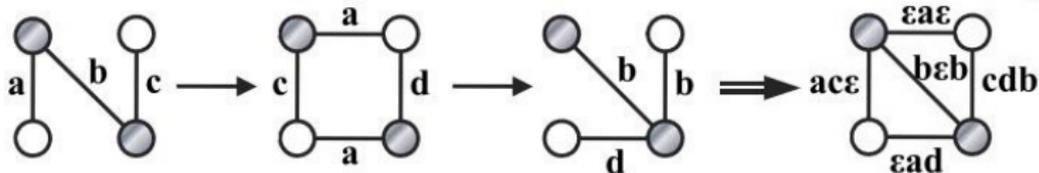
Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Transformação de um grafo no tempo em um grafo dinâmico. Os três grafos da esquerda são a representação de inserções e remoções de links no tempo. O grafo da direita é um grafo dinâmico que sumariza toda as informações que ocorreram ao longo do tempo.



Abordagens na descoberta de Sub-grafos

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- Muitas propostas para esta tarefa usam as propriedades *a priori* dos itens mais freqüentes do conjunto de dados.
- A geração de sub-estruturas é realizada numa fase de “matching”. É necessário gerar sub-grafos isomorfos.
- (Kuramochi et al.) usa uma representação de adjacência dos dados do grafo para melhorar a geração da sub-estrutura candidata, é aplicável para redes estáticas.
- (Wackersreuther et al.) baseado na idéia de árvores de sufixos encontra as sub-estruturas mais freqüentes mediante a união de grafos numa série de tempo.
- Técnicas de programação lógica indutiva são usadas para encontrar padrões freqüentes num domínio de toxicologia (Deshape et al.)



Descoberta e Compressão de Sub-grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- Uma proposta para a geração e compressão eficiente de sub-estrutura mais freqüentes é mediante o uso de busca com heurística greedy local (Matsuda et al.).
 - Subdue está baseada na heurística MDL (*Minimum Description Length*) para encontrar a estrutura padrão que permite uma melhor compressão do grafo (Cook and Holder).
 - GBI - Indução Baseada em Grafos, faz a compressão do grafo pela fragmentação de pares de vértices que aparecem com mais freqüência (Yoshida et al.).

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

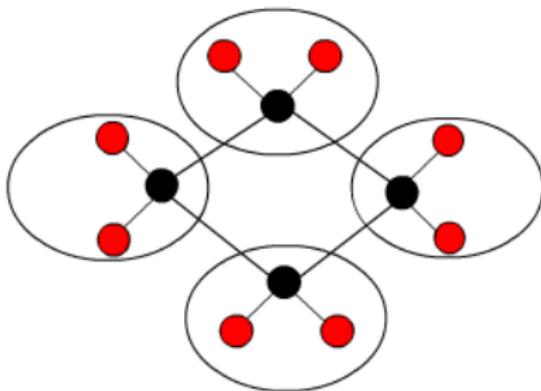
Descoberta de Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteccção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

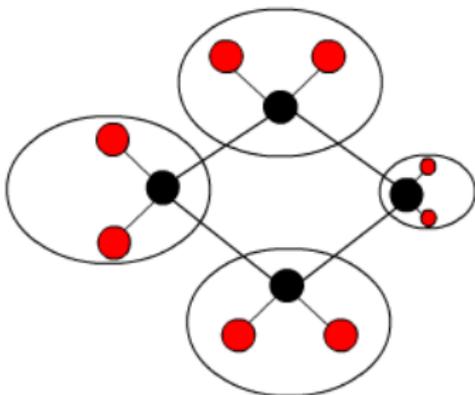
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteccção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

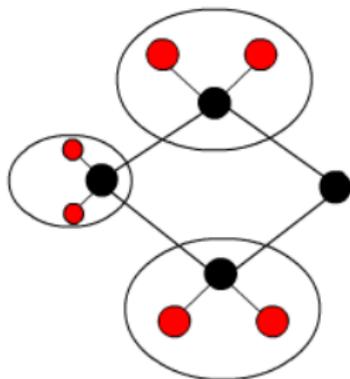
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteccão de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

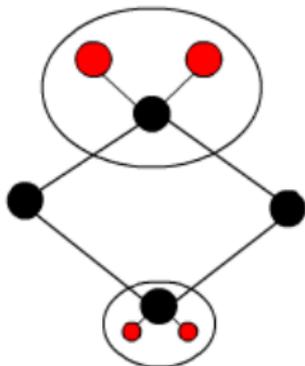
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

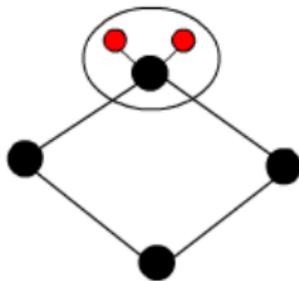
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteccão de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

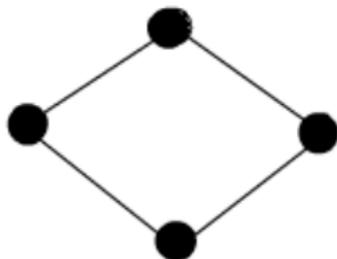
**Descoberta de
Sub-grafos**

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências





Classificação de Grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

**Classificação de
Grafos**

Geração de
Modelos para
grafos

Conclusões

Referências

- A classificação de grafos é um problema de aprendizado supervisionado que tem o objetivo de categorizar um grafo inteiro como uma instância positiva ou negativa de um conceito.
- A diferença com a tarefa de classificação de objetos baseada em links é que a classificação de grafos não precisa (sempre) da inferência coletiva, devido que os grafos são geralmente gerados independentemente.

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

**Classificação de
Grafos**

Geração de
Modelos para
grafos

Conclusões

Referências

- Três abordagens para a classificação de grafos foram exploradas:
 - 1 Mineração de Padrões de Grafos (*Feature Mining on Graphs*)
 - 2 Programação Lógica Indutiva (*Inductive Logic Programming*)
 - 3 Kernels de grafos (*Graph Kernels*)



Mineração de Padrões de Grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- Usa métodos relacionados à descoberta de sub-grafos.
- Seu processo geralmente começa com a descoberta de todas as sub-estruturas mais freqüentes ou informativas no grafo.
- As sub-estruturas descobertas são utilizadas para fazer a transformação dos dados do grafo em dados representados por uma tabela simples.
- Quando os grafos são representados por tabelas simples são utilizados os classificadores tradicionais para fazer a tarefa de classificação das instâncias na tabela.



Técnicas para a Mineração de Padrões de Grafos

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- A presença de grafos rotulados e não rotulados é um problema na classificação de grafos, (Kong and Yu) propor uma seleção de padrões semi-supervisionada e o uso de um algoritmo de ramificação e acotamento (*branch and bound*) para a busca eficiente das sub-estruturas ótimas do grafo e seu posterior classificação.
- A classificação de grafos concentra-se na colocação de uma única etiqueta para um grafo, (Kong and Yu 2) propor a colocação de muitas etiquetas simultaneamente para um grafo baseado na extração de bons padrões a partir do critério de independência de Hilbert-Schmidt e o uso de um algoritmo de ramificação e acotamento (*branch and bound*) para a busca eficiente das sub-estruturas ótimas do grafo.



Programação Lógica Indutiva

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- É um abordagem que utiliza sistemas de programação lógica indutiva para criar uma hipóteses que permita fazer a classificação.
- (King et. al) constrói um mapa dos dados do grafo que descreve a mutagênese na representação relacional dos dados. Seu representação lógica utiliza relações dos vértices com as arestas para depois utilizar um sistema de programação lógica indutiva para a busca de uma hipótese em seu espaço.



Kernels de Grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- Encontrar todas as sub-estruturas de um grafo numa coleção de grafos é computacionalmente proibido. Uma alternativa é o uso de kernels.
- (Gartner) propor um kernel que quantifica os caminhos que têm rótulos iguais no começo e no final.
- (Kashima and Inokuchi) propor um kernel que obtem a probabilidade dos caminhos aleatórios com iguais seqüências dos rótulos.



Geração de Modelos para grafos

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

**Geração de
Modelos para
grafos**

Conclusões

Referências

- É a tarefa de mineração de links relacionada com grafos que tenta desenvolver métodos de construção de modelos sobre conjuntos de dados com estruturas de grafos.
- A partir de um conjunto de grafos tenta gerar novos grafos que fazem parte da distribuição do conjunto de grafos original.
- Os modelos de geração para uma grande gama de tipos de grafos têm sido estudadas extensivamente na comunidade de análise de redes sociais.



Abordagens na geração de modelos para grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

**Geração de
Modelos para
grafos**

Conclusões

Referências

- Os grafos de Bernoulli (também conhecido como modelo Erdos-Rényi ou grafos aleatórios) propostos em (Frank and Nowicki), assume que as arestas dirigidas que pertencem a os objetos origem e destino têm uma distribuição idêntica e independente (IID) e dessa maneira estabelece que quando a probabilidade da existência de um link é 0.5 se tem um grafo com distribuição aleatória uniforme.



Abordagens na geração de modelos para grafos

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

**Geração de
Modelos para
grafos**

Conclusões

Referências

- Em (Kubika et al.) é apresentado um modelo de geração para links observados entre os indivíduos dada sua participação na rede que pertencem.
- Em (Kubika et al. 2) apresenta-se um modelo de geração de links para uma análise de links e consultas de colaboração que permitem tipos diferentes de links e informação temporal.
- Em (Getoor et al. 2) apresenta-se um modelo relacional probabilístico que fornece um modelo de geração unificado para objetos e links.



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Conclusões

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- As tarefas realizadas nos grafos baseadas em seus links, trabalham numa coleção de grafos e consideram um grafo como uma entidade relacionada cujos atributos são gerados pelos padrões existentes no mesmo grafo (relações entre vértices e atributos dos vértices relacionados).
- A tarefa de descoberta de sub-grafos é importante porque encontra sub-estruturas freqüentes, que por sua vez, são importantes no grafo, estas sub-estruturas permitem a compressão do grafo o qual é necessário para redes muito grandes e que apresentam muitas sub-estruturas como por exemplo nas redes de biologia ou química.

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- A tarefa de classificação de grafos permite rotular um grafo numa coleção de grafos, a capacidade de atribuir vários rótulos a um único grafo é umas dos desafios que tem esta tarefa.
- A tarefa de geração de grafos permite gerar um novo grafo a partir de um conjunto de grafos, sendo que o novo grafo pode apresentar as melhores estruturas do conjunto de grafos original. O uso das informações dos links é muito importante nesta tarefa porque permite a geração de um modelo com estruturas de dependência mais gerais.

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- A tarefa de detecção de grupo além da medição das características dos nós, também obtém informações das características de suas ligações e suas frequências no tempo pelo que resulta um método muito eficiente para encontrar grupos ocultos numa rede real.
- A tarefa de identificação de entidade tem a capacidade de fornecer uma medição muito precisa para identificar duplicações de dados e identifica-los, a diferença de outras medições que não tem em conta a hierarquia o ligações, senão só as características dos objetos.
- A tarefa de predição de ligações é muito difícil, porque a probabilidade a priori de uma ligação, é muito pequena e afeta o nível de confiança das predições



Conclusões

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- A tarefa de predição de ligações é aplicável aos problemas em que as ligações são desconhecidas, e contribui ao desenvolvimento de outras áreas, como a biologia.
- A tarefa de predição de ligações pode ser aplicada também para obter informações de eventos futuros, com as interações entre os membros de uma rede social ou redes de co-autoria.



Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Referências



Referências

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- (Deshaspe et al.) L. Dehaspe, H. Toivonen and R. King, Finding Frequent Substructures in Chemical Compounds, Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), 1998.
- (Kuramochi et al.) M. Kuramochi and G. Karypis. Frequent subgraph discovery. In IEEE International Conference on Data Mining, pp. 313320, 2001.
- (Wackersreuther et al.) B. Wackersreuther, P. Wackersreuther, A. Oswald, C. Bóhm and K. Borgwardt. Frequent Subgraph Discovery in Dynamic Networks, MLG's 10, Washington pp. 155-162, 2010.
- (Matsuda et al.) T. Matsuda, T. Horiuchi, H. Motoda, and T. Washio. Extension of graph-based induction for general graph structured data. In PAKDD, pages 420431, 2000.



Referências

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo

Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

- (Cook and Holder) D. J. Cook and L. B. Holder. Substructure discovery using minimum description length and background knowledge. *Journal of Artificial Intelligence Research*, 1:231255, 1994.
- (Yoshida et al.) K. Yoshida, H. Motoda, and N. Indurkha. Graph based induction as a unified learning framework. *Journal of Applied Intelligence*, 4(3):297316, July 1994.
- (Kong and Yu) Kong, Xiangnan and Yu, P. S.. Semi-supervised feature selection for graph classification. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'10*, 793-802, 2010.
- (Kong and Yu 2) Kong, Xiangnan and Yu, Philip S.. Multi-label Feature Selection for Graph Classification. *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 2010.

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos
Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

- (King et. al) R. D. King, S. H. Muggleton, A. Srinivasan, and M. J. E. Sternberg. Structure-activity relationships derived by machine learning: The use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. National Academy of Sciences, 93(1):438442, 1996.
- (Gartner) T. Gartner. Exponential and geometric kernels for graphs. In NIPS Workshop on Unreal Data: Principles of Modeling Nonvectorial Data, 2002.
- (Kashima and Inokuchi) H. Kashima and A. Inokuchi. Kernels for graph classification. In ICDM Workshop on Active Mining, 2002.
- (Getoor et. al) L. Getoor and C.P. Diehl. Link mining: a survey. SIGKDD Explor. Newsl., 7:3-12, December 2005.
- (Getoor et al. 2) L. Getoor, N. Friedman, D. Koller, and B. Taskar (2003). Learning probabilistic models of link structure. Journal of Machine Learning Research, 3:679707.



Referências

Introdução

Tarefas com Objetos - Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com links

Predição de
links

Tarefas com Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

- (Adibi J. et. al) J. Adibi, H. Chalupsky, E. Melz, and A. Valente. The Kojak group finder: Connecting the dots via integrated knowledge-based and statistical reasoning. In Proceedings of the Sixteenth Innovative Applications of Artificial Intelligence Conference IAAI-04, 800-807, 2004.
- (Ozgul et. al) F. Ozgul, J. Bondy, and H. Aksoy. Mining for offender group detection and story of a police operation. In Proceedings of the sixth Australasian conference on Data mining and analytics - Volume 70, AusDM 2007m 189-193, Darlinghurst, Australia, 2007. Australian Computer Society Inc.
- (Liben-Nowell and Kleinberg) D. Liben-Nowell and J. Kleinberg. The Link Prediction Problem for Social Networks. In Proceeding 11th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining. 2005.



Referências

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

- (Wasserman and Faust) S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications (Structural Anaylysis in the Social Sciences), p.249. Cambridge University Press, thirteenth edition, 1994.
- (Ananthakrishna et. al.) R. Ananthakrishna, S. Chaudhuri, and V. Ganti. Eliminating fuzzy duplicates in data warehouses. In International Conference on Very Large Databases (VLDB), Hong Kong, China, 2002.
- (Lü and Zhou) L. Lü and T. Zhou. Role of weak ties in link prediction of complex networks. In Proceeding of the 1st ACM International workshop on Complex networks meet information and knowledge management. 2009.
- (Frank and Nowicki) O. Frank and K. Nowicki (1993). Exploratory statistical analysis of networks. Annals of Discrete Mathematics, 55:349366.



Referências

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Deteção de
grupo
Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos
Classificação de
Grafos
Geração de
Modelos para
grafos

Conclusões

Referências

- (Song et al.) H. H. Song, T. Cho, V. Dave, Y. Zhang, and L. Qiu. Scalable proximity estimation and link prediction in online social networks. In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference. 2009.
- (Kubika et al.) J. Kubica, A. Moore, J. Schneider, and Y. Yang (2002). Stochastic link and group detection. In Eighteenth National Conference on Artificial Intelligence, pp. 798804. American Association for Artificial Intelligence.
- (Kubika et al. 2) J. Kubica, A. Moore, D. Cohn, and J. Schneider (2003). cGraph: A fast graph-based method for link analysis and queries. In IJCAI 2003 Text-Mining and Link-Analysis Workshop.



Perguntas?

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Perguntas?



Obrigado

Introdução

Tarefas com
Objetos -
Vértices

Links Baseados
em Ranking de
Objetos

Links Baseados
em Classificação
de Objetos

Detecção de
grupo

Resolução de
entidade

Tarefas com
links

Predição de
links

Tarefas com
Grafos

Descoberta de
Sub-grafos

Classificação de
Grafos

Geração de
Modelos para
grafos

Conclusões

Referências

Obrigado

