

Análise de Redes Sociais

Predição de Relacionamentos

Prof. Luciano Antonio Digiampietri
EACH-USP

Predição de Relacionamentos

Um dos aspectos mais estudados na ARS em relação à dinâmica das redes é a predição de relacionamentos (*Link Prediction*).

Predição de Relacionamentos

A predição de relacionamentos visa a identificar/sugerir relacionamentos que não estão presentes na rede social.

Pode ser focada em um dos três subproblemas apresentados a seguir.

Predição de Relacionamentos

A predição de relacionamentos visa a identificar/sugerir relacionamentos que não estão presentes na rede social.

1) Identificação de Relacionamentos Faltantes

Predição de Relacionamentos

A predição de relacionamentos visa a identificar/sugerir relacionamentos que não estão presentes na rede social.

- 1) Identificação de Relacionamentos Faltantes
- 2) Predição de Relacionamentos Inéditos

Predição de Relacionamentos

A predição de relacionamentos visa a identificar/sugerir relacionamentos que não estão presentes na rede social.

- 1) Identificação de Relacionamentos Faltantes
- 2) Predição de Relacionamentos Inéditos
- 3) Predição de Relacionamentos Futuros (Inéditos e Reincidentes)

Predição de Relacionamentos

Identificação de Relacionamentos Faltantes
(*Missing Link Prediction*):

Predição de Relacionamentos

Identificação de Relacionamentos Faltantes
(*Missing Link Prediction*):

Identificar relacionamentos que estão presentes na “vida real”, mas ausentes nas redes sociais.

Predição de Relacionamentos

Identificação de Relacionamentos Faltantes (*Missing Link Prediction*):

Identificar relacionamentos que estão presentes na “vida real”, mas ausentes nas redes sociais.

Duas pessoas são amigas, colegas de trabalho etc., mas este relacionamento não está presente na rede social.

Predição de Relacionamentos

Predição de Relacionamentos Novos/Inéditos:

Predição de Relacionamentos

Predição de Relacionamentos Novos/Inéditos:

Identificar novos relacionamentos (que potencialmente não existem ainda nem na “vida real”).

Predição de Relacionamentos

Predição de Relacionamentos Novos/Inéditos:

Identificar novos relacionamentos (que potencialmente não existem ainda nem na “vida real”).

Identificar quais pesquisadores (que nunca publicaram juntos) irão publicar no futuro (em uma rede de coautorias científicas).

Predição de Relacionamentos

Predição de Relacionamentos Futuros:

Predição de Relacionamentos

Predição de Relacionamentos Futuros:

Em uma rede social com informação temporal, prever quais serão os relacionamentos que ocorrerão num certo período de tempo futuro (novos e reincidentes)

Predição de Relacionamentos

Predição de Relacionamentos Futuros:

Em uma rede social com informação temporal, predizer quais serão os relacionamentos que ocorrerão num certo período de tempo futuro (novos e reincidentes)

Identificar quais pesquisadores irão publicar juntos no ano que vem (em uma rede de coautorias científicas).

**Mas como prever um
relacionamento?**

Predição de Relacionamentos



DIGI ♥ Match



?



Predição de Relacionamentos

E ACH_e seu Chamego



?



Aspectos Sociais Relevantes¹

¹Yin et al. (2010), Maruyama (2016)

Aspectos Sociais Relevantes¹

1. Homofilia: quanto mais interesses em comum as pessoas possuem, como características ou preferências, maior as chances de se relacionarem.

¹Yin et al. (2010), Maruyama (2016)

Aspectos Sociais Relevantes

2. Raridade: há mais chances de relacionamento entre pessoas com características ou preferências raras em comum. Pois características difíceis de encontrar tendem a se destacar em relação às outras características.

Aspectos Sociais Relevantes

3. Influência Social: uma característica compartilhada com muitos amigos de uma determinada pessoa pode ser útil para encontrar potenciais relacionamentos.

Aspectos Sociais Relevantes

3. Influência Social: uma característica compartilhada com muitos amigos de uma determinada pessoa pode ser útil para encontrar potenciais relacionamentos.

4. Amizades em Comum (ou vizinhos em comum): quanto mais amigos em comum duas pessoas possuem, maiores são as chances de se relacionarem.

Aspectos Sociais Relevantes

5. Proximidade Social: pessoas localizadas próximas em um grafo social possuem um relacionamento em potencial.

Aspectos Sociais Relevantes

5. Proximidade Social: pessoas localizadas próximas em um grafo social possuem um relacionamento em potencial.

6. Conexão Preferencial: pessoas populares tendem a atrair mais pessoas quando comparadas às pessoas com poucos relacionamentos.

Predição de Relacionamentos

A predição de relacionamentos pode ser vista e avaliada como um problema de:

Predição de Relacionamentos

A predição de relacionamentos pode ser vista e avaliada como um problema de:

Recomendação: para cada pessoa é preparada uma lista ordenada das principais sugestões.

Predição de Relacionamentos

A predição de relacionamentos pode ser vista e avaliada como um problema de:

Recomendação: para cada pessoa é preparada uma lista ordenada das principais sugestões.

Classificação Binária: “cada par” de pessoas é classificado como: terão ou não um relacionamento.

Predição de Relacionamentos

A predição de relacionamentos, tipicamente, é baseada em três tipos de métodos:

Predição de Relacionamentos

A predição de relacionamentos, tipicamente, é baseada em três tipos de métodos:

Baseados na Topologia da Rede: utilizam a estrutura da rede (por exemplo, número de vizinhos em comum).

Predição de Relacionamentos

A predição de relacionamentos, tipicamente, é baseada em três tipos de métodos:

Baseados nas Características dos Indivíduos:
recomendação de indivíduos com um mesmo perfil (sócio/econômico/geográfico), mesmos interesses em comum etc.

Predição de Relacionamentos

A predição de relacionamentos, tipicamente, é baseada em três tipos de métodos:

Baseados nas Características dos Indivíduos:
recomendação de indivíduos com um mesmo perfil (sócio/econômico/geográfico), mesmos interesses em comum etc.

Métodos Mistos: métodos que combinam topologia e características dos indivíduos.

Predição de Relacionamentos

Estratégias:

Recomendar de acordo com uma métrica.

Predição de Relacionamentos

Estratégias:

Recomendar de acordo com uma métrica.

Recomendar de acordo com uma combinação de métricas (regressão).

Predição de Relacionamentos

Estratégias:

Recomendar de acordo com uma métrica.

Recomendar de acordo com uma combinação de métricas (regressão).

Predizer de acordo com um conjunto de métricas (classificação).

Predição de Relacionamentos

A predição de relacionamentos utilizando aprendizado supervisionado (regressão ou classificação) precisa se atentar à seguinte questão:

Predição de Relacionamentos

A predição de relacionamentos utilizando aprendizado supervisionado (regressão ou classificação) precisa se atentar à seguinte questão:

Conjuntos de dados formados por todos os pares de indivíduos serão bastante desbalanceados para muitos problemas reais.

Predição de Relacionamentos

A predição de relacionamentos utilizando aprendizado supervisionado (regressão ou classificação) precisa se atentar à seguinte questão:

Conjuntos de dados formados por todos os pares de indivíduos serão bastante desbalanceados para muitos problemas reais.

Isto é, haverá muito mais pares que não se relacionarão do que o contrário.

(Algumas) Medidas Topológicas

(Algumas) Medidas Topológicas

Vizinhos em Comum (*Common Neighbors* - *CN*):

$$CN_{x,y} = |\Gamma(x) \cap \Gamma(y)|$$

sendo $|\Gamma(y)|$ o número de vizinhos de y

(Algumas) Medidas Topológicas

Vizinhos em Comum (*Common Neighbors - CN*):

$$CN_{x,y} = |\Gamma(x) \cap \Gamma(y)|$$

sendo $|\Gamma(y)|$ o número de vizinhos de y

Quanto maior o número de vizinhos em comum, maior a chance da existência de um relacionamento.

(Algumas) Medidas Topológicas

Coeficiente de Jaccard (*Jaccard Coefficient* - *JC*):

$$JC_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

(Algumas) Medidas Topológicas

Coeficiente de Jaccard (*Jaccard Coefficient* - *JC*):

$$JC_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

Vizinhos em Comum dividido/normalizado pelo número total de amigos.

(Algumas) Medidas Topológicas

Adamic-Adar (AA):

$$AA_{x,y} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(|\Gamma z|)}$$

(Algumas) Medidas Topológicas

Adamic-Adar (AA):

$$AA_{x,y} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(|\Gamma z|)}$$

Métrica que pondera a vizinhança em comum entre os vértices (vizinhos em comum com menos vizinhos pesam mais).

(Algumas) Medidas Topológicas

Alocação de Recursos (*Resource Allocation* - *AA*):

$$RA_{x,y} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{|\Gamma(z)|}$$

(Algumas) Medidas Topológicas

Alocação de Recursos (*Resource Allocation - AA*):

$$RA_{x,y} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{|\Gamma(z)|}$$

Métrica para simular a transmissão de recursos, assumindo que o recurso é dividido igualmente entre todos os vizinhos de cada indivíduo.

(Algumas) Medidas Topológicas

Salton ou Similaridade Cosseno (SA):

$$SA_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{|\Gamma(x)| * |\Gamma(y)|}}$$

(Algumas) Medidas Topológicas

Salton ou Similaridade Cosseno (SA):

$$SA_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{|\Gamma(x)| * |\Gamma(y)|}}$$

Medida de similaridade entre vetores adaptada para previsão de relacionamentos.

(Algumas) Medidas Topológicas

Leicht-Holme-Newman (*LHN*):

$$LHN_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x)| * |\Gamma(y)|}$$

(Algumas) Medidas Topológicas

Leicht-Holme-Newman (*LHN*):

$$LHN_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x)| * |\Gamma(y)|}$$

Mesma lógica da medida de Salton, mas com mais peso no denominador.

(Algumas) Medidas Topológicas

Coeficiente de Sørensen-Dice (*Sørensen-Dice Coefficient - SD*):

$$SD_{x,y} = \frac{2 * |\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x)| + |\Gamma(y)|}$$

(Algumas) Medidas Topológicas

Coeficiente de Sørensen-Dice (*Sørensen-Dice Coefficient - SD*):

$$SD_{x,y} = \frac{2 * |\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x)| + |\Gamma(y)|}$$

Medida utilizada inicialmente para medir similaridade entre amostras de dados de comunidades ecológicas.

(Algumas) Medidas Topológicas

Hub Promoted Index - HPI:

$$HPI_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\min(|\Gamma(x)|, |\Gamma(y)|)}$$

(Algumas) Medidas Topológicas

Hub Promoted Index - HPI:

$$HPI_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\min(|\Gamma(x)|, |\Gamma(y)|)}$$

Privilegia a interseção quando um dos indivíduos possui poucos amigos.

(Algumas) Medidas Topológicas

Hub Depressed Index - HDI:

$$HDI_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\max(|\Gamma(x)|, |\Gamma(y)|)}$$

(Algumas) Medidas Topológicas

Hub Depressed Index - HDI:

$$HDI_{x,y} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\max(|\Gamma(x)|, |\Gamma(y)|)}$$

Penaliza a interseção quando um dos indivíduos possui muitos amigos.

(Algumas) Medidas Topológicas

Conexão Preferencial (*Preferential Attachment* - *PA*):

$$PA_{x,y} = |\Gamma(x)| * |\Gamma(y)|$$

(Algumas) Medidas Topológicas

Conexão Preferencial (*Preferential Attachment - PA*):

$$PA_{x,y} = |\Gamma(x)| * |\Gamma(y)|$$

Há mais chance de ocorrerem novas conexões entre pessoas que possuem mais conexões. “O rico fica mais rico.”

(Algumas) Medidas Topológicas

As medidas apresentadas até o momento são todas baseadas no número de vizinhos (dos dois nós em análise e/ou dos nós que estão na intersecção de suas vizinhanças).

(Algumas) Medidas Topológicas

As medidas apresentadas até o momento são todas baseadas no número de vizinhos (dos dois nós em análise e/ou dos nós que estão na intersecção de suas vizinhanças).

Há diversas outras métricas, como as baseadas em caminhos.

(Algumas) Medidas Topológicas

Caminho Mais Curto (*Shortest Path* - *SP*):

$$SP_{x,y} = -dist(x, y)$$

**sendo *dist* a distância do menor caminho
(geodésico) de x a y**

(Algumas) Medidas Topológicas

Caminho Mais Curto (*Shortest Path - SP*):

$$SP_{x,y} = -dist(x, y)$$

sendo *dist* a distância do menor caminho
(geodésico) de x a y

**Quanto mais próximos os nós, maior a chance de
haver um relacionamento entre eles.**

(Algumas) Medidas Topológicas

Outras medidas baseadas em distância:

(Algumas) Medidas Topológicas

Outras medidas baseadas em distância:

Katz: realizada a soma de todos os caminhos entre dois vértices favorecendo os mais curtos.

(Algumas) Medidas Topológicas

Outras medidas baseadas em distância:

Katz: realizada a soma de todos os caminhos entre dois vértices favorecendo os mais curtos.

Hitting Time: número esperado de passos (usando caminhos aleatórios) para, partindo de um nó chegar no outro.

(Algumas) Medidas Topológicas

Outras medidas baseadas em distância:

Katz: realizada a soma de todos os caminhos entre dois vértices favorecendo os mais curtos.

Hitting Time: número esperado de passos (usando caminhos aleatórios) para, partindo de um nó chegar no outro.

Rooted Pagerank: usa a lógica de distribuição de prestígio do Pagerank. Quanto mais distantes os nós forem, menor será o prestígio recebido.

Características dos Indivíduos

Medidas baseadas nas características dos indivíduos dependem muito do contexto da rede social e das informações (de perfil ou de utilização da rede) que estão disponíveis.

Características dos Indivíduos

Por exemplo, no Facebook:

Características dos Indivíduos

Por exemplo, no Facebook:

Endereço;

Características dos Indivíduos

Por exemplo, no Facebook:

Endereço;

Educação (local ou curso);

Características dos Indivíduos

Por exemplo, no Facebook:

Endereço;

Educação (local ou curso);

Emprego (função ou local);

Características dos Indivíduos

Por exemplo, no Facebook:

Endereço;

Educação (local ou curso);

Emprego (função ou local);

Check-ins em comum;

Características dos Indivíduos

Por exemplo, no Facebook:

Endereço;

Educação (local ou curso);

Emprego (função ou local);

Check-ins em comum;

Hobbies;

Características dos Indivíduos

Por exemplo, no Facebook:

Endereço;

Educação (local ou curso);

Emprego (função ou local);

Check-ins em comum;

Hobbies;

Conteúdo das postagens;

Características dos Indivíduos

Por exemplo, no Facebook:

Endereço;

Educação (local ou curso);

Emprego (função ou local);

Check-ins em comum;

Hobbies;

Conteúdo das postagens;

Curtidas em páginas ou comentários ...

Características dos Indivíduos

Em redes sociais acadêmicas:

Características dos Indivíduos

Em redes sociais acadêmicas:

Endereço Profissional;

Características dos Indivíduos

Em redes sociais acadêmicas:

Endereço Profissional;

Áreas de Atuação;

Características dos Indivíduos

Em redes sociais acadêmicas:

Endereço Profissional;

Áreas de Atuação;

Participação de eventos;

Características dos Indivíduos

Em redes sociais acadêmicas:

Endereço Profissional;

Áreas de Atuação;

Participação de eventos;

Assunto das publicações;

Características dos Indivíduos

Em redes sociais acadêmicas:

Endereço Profissional;

Áreas de Atuação;

Participação de eventos;

Assunto das publicações;

Formação (local, época, área) ...

Características dos Indivíduos

Estas características podem ser usadas por classificadores ou regressores da mesma forma que as características estruturais.

Referências

- 1 MARUYAMA, William Takahiro. Predição de coautorias em redes sociais acadêmicas . 2016. Dissertação (Mestrado em Sistemas de Informação) - EACH-USP, São Paulo, 2016.
[doi:10.11606/D.100.2016.tde-07052016-232625](https://doi.org/10.11606/D.100.2016.tde-07052016-232625).
- 2 YIN, Z. et al. A unified framework for link recommendation using random walks. In: Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on., 2010. p. 152–159.
- 3 DAUD, Nur Nasuha et al. Applications of link prediction in social networks: A review. Journal of Network and Computer Applications, Volume 166, 2020.
- 4 WANG, P. et al. Link prediction in social networks: the state-of-the-art. Sci. China Inf. Sci. 58, 1–38 (2015).
<https://doi.org/10.1007/s11432-014-5237-y>

Análise de Redes Sociais

Predição de Relacionamentos

Prof. Luciano Antonio Digiampietri
EACH-USP