

Aprendizado Supervisionado Aplicado à Predição de Links

Celio Henrique Nogueira Larcher Junior

Laboratório Nacional de Computação Científica

Petrópolis, 2017

Agenda

- 1 O Problema de Predição de Links
- 2 Aprendizado Supervisionado Aplicado à Predição de Links
- 3 Experimentos Computacionais
- 4 Conclusões e trabalhos futuros
- 5 Referências

Agenda

- 1 O Problema de Predição de Links
- 2 Aprendizado Supervisionado Aplicado à Predição de Links
- 3 Experimentos Computacionais
- 4 Conclusões e trabalhos futuros
- 5 Referências

Predição de Links

- Definição: Dado um grafo variante no tempo $G(V, E, t)$, espera-se determinar as conexões que mais provavelmente de ocorrerão em um momento futuro
- Este momento futuro é, em geral, delimitado por uma janela de tempo Δt
- Desta forma, uma outra maneira de se referir ao problema é verificar a diferença entre os grafos $G(V, E, t)$ e $G(V, E, t + \Delta t)$

Aplicações

- Alguns exemplos de aplicações:
 - Sistemas de recomendação
 - Indicações para novas colaborações acadêmicas
 - Análise de possíveis conexões em redes de contato de terroristas

Agenda

- 1 O Problema de Predição de Links
- 2 **Aprendizado Supervisionado Aplicado à Predição de Links**
- 3 Experimentos Computacionais
- 4 Conclusões e trabalhos futuros
- 5 Referências

Aplicação na predição de links

- Algoritmos de aprendizado supervisionado buscam, através de exemplos, generalizar as relações presentes nos dados
- O problema de predição de links se adequa de forma interessante, dado que pode ser descrito como um problema de classificação com duas classes:
 - Classe 1: espera-se o surgimento de uma aresta entre o par de nós em dado intervalo de tempo
 - Classe 2: não é previsto o surgimento de aresta entre este par de nós neste intervalo de tempo
- Os dados de entrada do sistema de aprendizado são relativos às informações de cada par de nós

Objetivo

- Verificar quais atributos topológicos de um grafo são mais relevantes para a tarefa de predição de links via aprendizado supervisionado
 - Considerar diferentes modelos de grafo (GNP, SW, PA)
 - Verificação de diversas técnicas de aprendizado
 - Validar os resultados em grafos obtidos de aplicações reais

Agenda

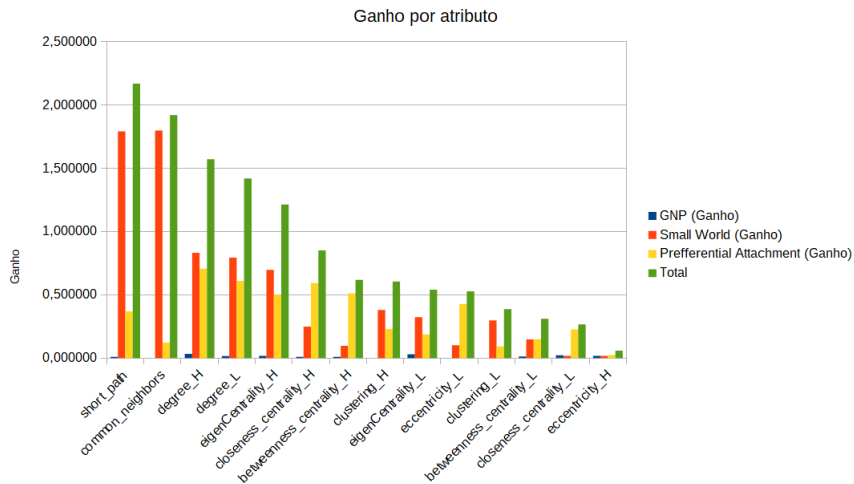
- 1 O Problema de Predição de Links
- 2 Aprendizado Supervisionado Aplicado à Predição de Links
- 3 Experimentos Computacionais**
- 4 Conclusões e trabalhos futuros
- 5 Referências

Experimento

- Experimentos realizados com auxílio do *framework* Weka
- Mediu-se a performance da classificação de cada atributo individualmente, em um conjunto de técnicas de aprendizado
- As técnicas de aprendizado selecionadas foram:
 - Árvore de Decisão (J48);
 - Random Forest;
 - Suport Vector Machine (SMO);
 - Naive Bayes;
- Como segunda etapa, mediu-se a performance da classificação no Top-5, comparado ao conjunto total de atributos

Atributos verificados

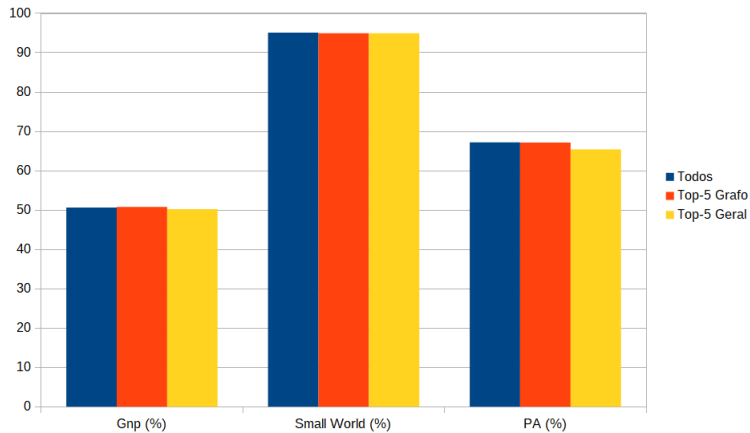
- *short_path*
- *common_neighbors*
- *degree_H*, *degree_L*
- *betweenness centrality_H*, *betweenness centrality_L*
- *closeness centrality_H*, *closeness centrality_L*
- *eigenCentrality_H*, *eigenCentrality_L*
- *clustering_H*, *clustering_L*
- *eccentricity_H*, *eccentricity_L*



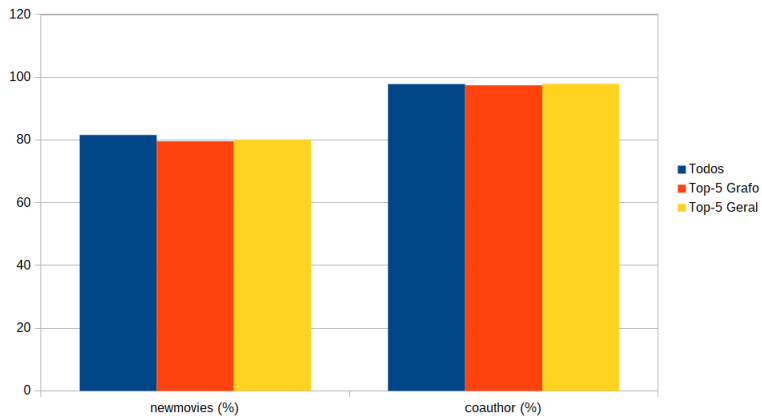
Atributos mais relevantes

- *short_path*
- *common_neighbors*
- *degree_H*
- *degree_L*
- *eigenCentrality_H*

Desempenho por conjunto de atributos



Desempenho por conjunto de atributos



Agenda

- 1 O Problema de Predição de Links
- 2 Aprendizado Supervisionado Aplicado à Predição de Links
- 3 Experimentos Computacionais
- 4 Conclusões e trabalhos futuros
- 5 Referências

Conclusões e trabalhos futuros

- Neste trabalho foi identificado um subconjunto de atributos significativo capaz de manter a capacidade de predição
- Estes atributos inclusive se apresentam na intersecção dos fatores motivadores de cada tipo de grafo analisado
- Outro fator, foi verificado um forte papel da política de estabelecimento das conexões na capacidade de predição
- Os experimento em grafos reais corroboram estas análises
- Como trabalhos futuros algumas possibilidades são a utilização de novas métricas e verificação dos resultados em grafos dinâmicos

Agenda

- 1 O Problema de Predição de Links
- 2 Aprendizado Supervisionado Aplicado à Predição de Links
- 3 Experimentos Computacionais
- 4 Conclusões e trabalhos futuros
- 5 Referências**

Referências I



Al Hasan, M., Chaoji, V., Salem, S., Zaki, M. (2006).

Link prediction using supervised learning. *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security.*



Benchettara, N., Kanawati, R., Rouveirol, C. (2010).

Supervised machine learning applied to link prediction in bipartite social networks. *Proceedings - 2010 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010*



Cukierski, W., Hamner, B., Yang, B. (2011).

Graph-based features for supervised link prediction. *Proceedings of the International Joint Conference on Neural Networks*



Humphries M. D., Gurney K. (2008).

Network “Small-World-Ness”: A Quantitative Method for Determining Canonical Network Equivalence. *Sporns O, ed. PLoS ONE*



Martinez, V., Berzal, F., Cubero, J. (2016).

A Survey of Link Prediction in Complex Networks. *ACM Computing Surveys*

Referências II



de Sa, H. R., Prudencio, R. B. C. (2011).

Supervised link prediction in weighted networks. *In The 2011 International Joint Conference on Neural Networks*



Tang, J., Sun, J., Wang, C., Yang, Z. (2009).

Social Influence Analysis in Large-scale Networks. *In Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*



Wang, L., Lou, T., Tang, J., Hopcroft, J. (2011).

Detecting Community Kernels in Large Social Networks. *Proceedings of 2011 IEEE International Conference on Data Mining.*



Witten, I. H., Frank, E. (2005).

Data Mining: Practical Machine Learning Tools and Techniques. *San Francisco, CA, USA: Morgan Kaufmann Publishers Inc*