

PREDIÇÃO DE LINKS EM REDES DE COAUTORIA: UM ESTUDO DA INFLUÊNCIA DAS CARACTERÍSTICAS DA REDE NA EFICIÊNCIA DOS ALGORITMOS

GOMES, Orlando Abreu, orlando.gomes@fumec.br; PARREIRAS, Fernando Silva, fernando.parreiras@fumec.br; DINIZ, Mateus Esdras Souza, linux.esdras@gmail.com
FACE, Universidade FUMEC, Belo Horizonte, MG

RESUMO

Rede de coautoria é uma rede social na qual os autores são os nós e os coautores de um artigo indicam as arestas dessa rede. Este trabalho buscou identificar as características de redes de coautoria que impactam na eficácia da predição de links feita por algoritmos baseados na topologia de quatro bases de dados. De acordo com as análises, os melhores resultados foram obtidos com o uso de redes de coautoria grandes, com baixa densidade e grande diâmetro, formadas por autores das ciências humanas. Em relação ao período de tempo utilizado, os melhores resultados foram obtidos com o período presente de dois anos e o período futuro de um trimestre.

INTRODUÇÃO

Redes estão em todos os lugares. Sempre que se observam interações entre objetos, uma rede é formada implicitamente. Redes surgem em quase todas as áreas de pesquisa, comércio e sociedade: redes sociais, redes de rodovias, redes de comunicação, redes de hiperlinks da web, redes neurais, redes de coautorias, na qual os autores são os vértices e os coautores de um artigo indicam as arestas, dentre outras. É natural supor que as redes evoluam no tempo. Considerando um instantâneo de uma rede de coautoria científica, por exemplo, a predição de links é uma ferramenta que possibilita, em um momento posterior a esse instantâneo, a recomendação de uma nova parceria entre cientistas dessa rede [1]. Prever conexões em redes sociais de coautoria contribui para o crescimento científico, facilitando a colaboração entre pesquisadores com potencial de contribuição mútua [2]. Este trabalho objetivou identificar as características das redes que impactam na eficácia da predição de links feita por algoritmos baseados na topologia.

METODOLOGIA

Inicialmente, realizou-se uma revisão da literatura a respeito de predição de links em redes de coautoria [3]. Obtiveram-se as redes de coautoria a serem investigadas. Foram elaborados programas na linguagem Python para realizar a predição de links nas redes de coautoria obtidas. Determinou-se a melhor técnica de mensurar os resultados obtidos por cada rede de coautoria. Caracterizaram-se as redes de coautoria escolhidas para correlacionar com os resultados obtidos por cada técnica de predição de links analisada. Finalmente, compararam-se os resultados obtidos a fim de determinar quais características das redes de coautoria contribuem de forma positiva ou negativa predição de links para cada técnica clássica de predição investigadas.

RESULTADOS E DISCUSSÃO

Foram analisadas 122 referências que permitiram concluir que não há uma investigação semelhante à proposta do projeto. Os algoritmos de predição de links escolhidos enquadram-se na categoria das métricas de topologia que são bastante populares por sua eficiência e simplicidade que foram as técnicas: vizinhos comuns, anexação preferencial, coeficiente de Jaccard e Adamic-Adar. Foram selecionadas quatro bases: a base de currículos Lattes, a base DBLP (Digital Bibliography & Library Project), ArXiv (uma pronúncia para a palavra "archive") e a MAG (Microsoft Academic Graph). É importante ressaltar que a obtenção das bases, bem como o seu tratamento para futuras análises, não é um procedimento trivial. Os critérios para selecionar os parâmetros a serem analisados basearam-se na viabilidade de obtê-los e, também, em critérios que já foram investigados em outros trabalhos que foram: tamanho da rede, período presente, densidade da rede, diâmetro da rede, período futuro, área de atuação e tipos de publicação. Nessa análise, foram realizados mais de 1.000 experimentos nas quais quatro bases de coautoria, citadas anteriormente, foram segmentadas e mapeadas para permitir a escolha de conjuntos similares que variassem somente a característica investigada.

CONCLUSÃO

De acordo com as análises, os melhores resultados foram obtidos com o uso de redes de coautoria grandes, com baixa densidade e grande diâmetro, formadas por autores das ciências humanas. Em relação ao período de tempo utilizado, os melhores resultados foram obtidos com o período presente de dois anos e o período futuro de um trimestre. As análises efetuadas neste trabalho ajudarão a compreender melhor a dinâmica das redes sociais acadêmicas e também contribuirão para a escolha dos melhores algoritmos de predição de link para cada tipo de rede social. A definição de critérios para a escolha de conjuntos que favoreçam a comparação de resultados entre diferentes pesquisas e algoritmos é outra contribuição relevante.

REFERÊNCIAS

- [1] LIBEN-NOWEL, D.; KLEINBERG, J. The Link-Prediction Problem for Social Networks. *Journal of the American Society for Information Science and Technology*, v. 58(7), p. 1019-1031, 2007.
- [2] PAVLOV, M; ICHISE, R. Finding Experts by Link Prediction in Co-authorship Networks. *2nd International ExpertFinder Workshop* (2007).
- [3] WANG, Peng et al. Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, [Beijing], v. 58, n. 1, p. 1-38, Jan. 2015.