



UNIVERSIDADE
FUMEC

DE MINAS GERAIS PARA O MUNDO

RECOMENDAÇÃO DE NOVAS COLABORAÇÕES ENTRE PESQUISADORES POR MEIO DO CURRÍCULO LATTES

GOMES, Orlando Abreu, orlando.gomes@fumec.br; PARREIRAS, Fernando Silva, fernando.parreiras@fumec.br; DIAS JÚNIOR, Alair, alair@cpdee.ufmg.br; CUNHA, Thiago Chierici, thiago.cunha@gmail.com; DA SILVA JUNIOR, Evando Alves, evando66.ea@gmail.com
FACE, Universidade FUMEC, Belo Horizonte, MG

RESUMO

Nesse projeto foram testados métodos de predição de links – vizinhos comuns, Adamic-Adar, Jaccard e anexação preferencial – para recomendação de novas colaborações entre pesquisadores doutores na plataforma de currículos Lattes. A ligação entre os pesquisadores da rede foi determinada pela coautoria em artigos que se encontram no currículo Lattes de cada pesquisador doutor. Foram analisadas, na plataforma Lattes, as bases de dados das Ciências Exatas, Engenharias e bolsistas de produtividade CNPq pelos métodos citados anteriormente. O método de predição de links Adamic-Adar foi o que apresentou o melhor desempenho para os três subconjuntos escolhidos – 1.449 vezes superior ao aleatório.

INTRODUÇÃO

Redes estão em todos os lugares. Uma rede de coautoria se forma quando os autores de um artigo são os vértices e os coautores do mesmo indicam as arestas. 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]. A plataforma Lattes constitui-se em uma base de currículos de pesquisadores, professores e estudantes do Brasil. A análise dessa base de dados pode permitir identificar potenciais grupos de cooperação [2]. Portanto, um método que permita a predição de links em uma rede de coautoria pode ser de grande valia para incrementar a produção científica do país. Neste trabalho, objetiva-se testar e determinar, dentre os métodos de predição de links clássicos na literatura, o melhor algoritmo para predição de links na rede de coautoria da plataforma Lattes.

METODOLOGIA

Para alcançar a meta, realizou-se um estudo das técnicas clássicas de predição de links clássicos: o método dos vizinhos comuns, o método Adamic-Adar, Jaccard e o método de anexação preferencial [1]. Para cada uma destas técnicas foi escrito um programa na linguagem Python para analisar uma sub-rede da plataforma de coautoria Lattes. Foram analisadas, na plataforma Lattes, as bases de dados das Ciências Exatas, Engenharias e bolsistas de produtividade CNPq pelos métodos citados anteriormente. A razão de se escolher subgrupos da rede de coautoria da plataforma Lattes, foi limitação de memória e velocidade de processamento.

RESULTADOS E DISCUSSÃO

Foram analisadas, na plataforma Lattes, as bases de dados das Ciências Exatas, Engenharias e bolsistas de produtividade CNPq pelos seguintes métodos de predição de links: vizinhos comuns (VC), Adamic-Adar (AA), Jaccard (JC) e anexação preferencial (AP). Utilizou-se a base de dados até o ano de 2012 e realizou-se a predição de links para o ano de 2013. Os resultados são apresentados na Tabela I. Percebe-se que o método de predição de links Adamic-Adar é o que apresenta o melhor desempenho para os três subconjuntos escolhidos. Em relação ao processo aleatório, o processo se apresenta, em média, 1.449 vezes superior ao aleatório.

SUBGRUPO	ALEATÓRIO (%)	JACCARD (%)	ANEXAÇÃO PEF. (%)	ADAMIC-ADAR (%)	VIZINHOS COMUNS (%)
EXATAS	0,0016	3,7581	0,2099	5,0875	4,2579
PRODUTIVID.	0,0079	5,0299	0,2661	6,4538	5,3094
ENGENHARIA	0,0014	3,1647	0,3120	4,2567	3,2538
MÉDIA	0,0036	3,9843	0,2627	5,2660	4,2737
MÉDIA/ALEATÓRIO	1,00	1.097	72	1.449	1.176

Tabela I – Percentual de acerto de cada método de predição de links para subconjuntos da plataforma de currículos Lattes.

CONCLUSÃO

Conseguiu-se realizar a predição de links em subgrupos da plataforma de currículos Lattes, utilizando-se os métodos clássicos da literatura. O melhor algoritmo de recomendação de links foi o Adamic-Adar no qual se obteve, em média, o percentual de acerto de 5,3%. Apesar de não ser um valor encorajador, o algoritmo é 1.449 vezes superior ao aleatório. Em trabalhos futuros sugere-se procurar outros métodos de predição de links mais eficientes [3].

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] LUNA, J.E.O.; REVOREDO, K.; COZMAN, F.G. Link prediction using a probabilistic description logic. *Journal of the Brazilian Computer Society*, v. 19(4), p. 397-409.