

TFG I - Proposta

Sistema de Recomendação Baseado em Confiança para
Promover a Colaboração em Redes de Pesquisa Científica

J.P. Saldanha¹ F. Prass¹

¹Ciência da Computação
Universidade Franciscana

Jul. 2019 / TFG I

Resumo

- Sistema de Recomendação
 - Promover Colorações
- Usuários
 - Pesquisadores
- Metodologia
 - Método Baseado em Conteúdo
 - Método Baseado em Confiança
- Dados
 - Plataforma Lattes
 - Plataforma Kennis

Pesquisa Científica

- Incrementar o Conhecimento
 - Referências
 - Colaborações
- Não reinventar a roda
- "Stand on the Shoulders of Giants" - Google Acadêmico

Plataforma Lattes

- Base de dados de Currículos

CNPq

"Padrão nacional no registro da vida pregressa e atual dos estudantes e pesquisadores do país"

Sobrecarga de Informações

- Em um ano...
 - 2.5 milhões de artigos por ano
 - Aumento de 5% no número de autores
- Impossível considerar
 - Relacionados
 - Linhas de pesquisa
 - Pesquisadores
- Trabalho repetido
- Perda da oportunidade de colaborações

Sistemas de Recomendação

- Pessoas tendem a considerar recomendações de amigos/especialistas
- Analisar tendências dentro de comunidades
- Filtragem das informações/itens em websites

Objetivos

Geral

Promover a colaboração de pesquisadores em trabalhos científicos

- Estudar funções e aplicações de sistemas de recomendação
- Modelar a rede de confiança da comunidade científica
- Estabelecer métricas de confiança para os dados disponíveis
- Estimar a confiança entre os pesquisadores
- Pré-selecionar recomendações
- Filtrar a pré-seleção com a confiança computada
- Avaliar o modelo

Proposta

- Descrever uma rede de colaborações
- Computar a propagação de confiança na rede
- Pré-seleção dos itens com o método baseado em conteúdo
- Arquitetura para o sistema de recomendações

Sistemas de recomendação (SR)

- Filtrar a grande quantidade de itens das plataformas
- Predizer a relevância do Item
- Usuários
 - $u_1, \dots, u_n \in U$
- Itens
 - $i_1, \dots, i_n \in I$
- Relações
 - Ontologia
 - Matriz $|U| \times |I|$
- Função de predição
 - Real: $f(u, i)$
 - Computada: $\hat{f}(u, i)$

Tipos de SR

Filtragem Colaborativa

Itens bem avaliados pelos vizinhos serão também avaliados positivamente pelo usuário alvo, já que os perfis são semelhantes

Baseado em Conteúdo

Usuários têm interesse em itens semelhantes àqueles que lhe foram uteis no passado

Baseado em Confiança

O usuário recebe recomendações de itens avaliados positivamente por usuários em sua rede de confiança

Desenvolvimento de um Sistema de Recomendação para Bibliotecas Digitais

Furlan et al. 2018

- Aborda o problema da sobrecarga de informações
 - Baseando-se no perfil do currículo Lattes
- Aborda o problema da sobrecarga de informações
 - Filtragem colaborativa & Baseado em conhecimento
 - Problema da primeira avaliação

Técnicas de Recomendação para usuários de Bibliotecas Digitais

Primo & Loh 2006

- Apresentam técnicas populares de recomendação
- Abordagens para a elaboração de um SR de obras literárias
- SR Híbrido
- Comparativo das metodologias usadas

Trust-aware Collaborative Filtering for Recommender Systems

Massa & Avesani 2004

- Melhorar sugestões em SRs usando métricas de confiança
- Distância mínima entre nós
- PageRank como métrica global
- Arquitetura com módulos substituíveis

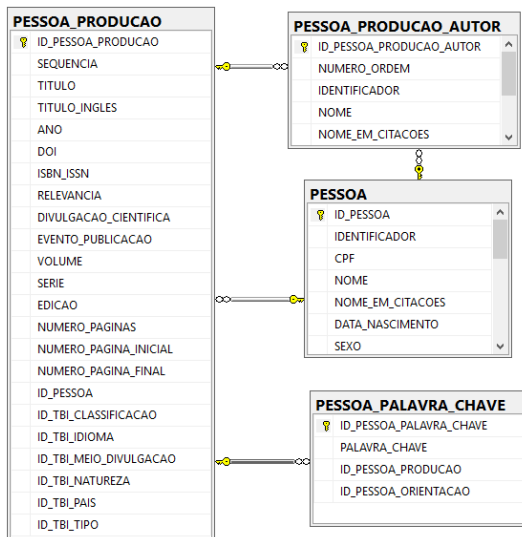
Materiais & Métodos

- Estimar a confiança entre pesquisadores
 - Modelar uma rede de confiança
 - Local ou global
- Selecionar potenciais colaboradores baseando-se no perfil
- Ordenada de acordo com a confiança estimada

Os Dados da Plataforma Lattes

- Plataforma Kennis
 - Parser do currículo Lattes
 - Limpeza e pré-processamento
- Banco de dados relacional

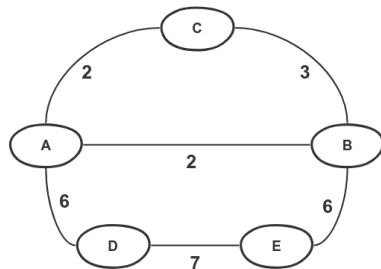
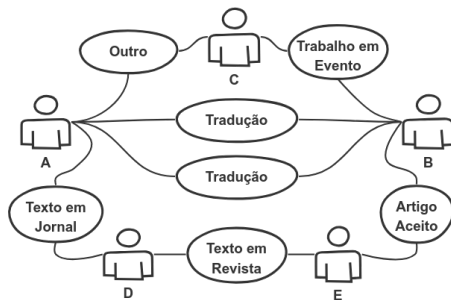
Diagrama ER



Rede de Publicações

- Rede de confiança
 - Nós: Pesquisadores
 - Arestas: publicações em conjunto
- Propagação & Agregação de confiança

Rede de Confiança



PageRank

Page et al. 1999

- Ranking global de citações
- Dissipação de ranking
 - Seja R o ranking de uma página qualquer, u
 - R é a soma dos rankings dissipados pelas páginas que referenciam u (B_u)
 - R é distribuído igualmente entre as páginas as quais u faz referência (F_u)

PageRank

■ Matriz de adjacência

$$\hat{A}_{u,v} = \begin{cases} 1 & \text{se há links de } u \text{ para } v \\ 0 & \text{se não há links de } u \text{ para } v \end{cases} \quad (1)$$

$$\blacksquare A = \frac{\hat{A}}{|F|}$$

$$\blacksquare R = c(AR + E)$$

PageRank

- *"A aplicação de PageRank à um grafo não-direcionado gera um vetor R estatisticamente similar à distribuição de grau dos nodos da rede"*
- No final das contas a confiança seria proporcional ao número de publicações do autor
- Não é considerada a confiança dos colaboradores, somente o valor total de colaborações

Relevância

Tipo de Publicação	Peso
Livro Publicado ou Organizado	9
Capítulo de Livro Publicado	8
Artigo Publicado	8
Artigo Aceito para Publicação	7
Texto em Jornal/Revista	6
Trabalho em Evento	3
Outra Produção Bibliográfica	2
Prefácio/Posfácio	2
Tradução	1

Centralidade

Opsahl et al. 2010

- Seguir a ideia de *PageRank*
- Cálculo da centralidade dos nodos.
- Relevância & quantidade de colaborações devem ser levadas em consideração na distribuição de confiança.

Centralidade

Opsahl et al. 2010

$$C_D^{w\alpha}(i) = k_i^{1-\alpha} \times s_i^\alpha \quad (2)$$

Nodo	$C_D^{w\alpha}(i)$			
	$\alpha = 0.00$	$\alpha = 0.50$	$\alpha = 1.00$	$\alpha = 1.50$
A	3.00	5.47	10.00	18.25
B	3.00	5.74	11.00	21.06
C	2.00	3.16	5.00	7.90
D	2.00	5.09	13.00	33.00
E	2.00	5.09	13.00	33.00

Distância

Opsahl et al. 2010

- Confiança local
- Distância entre os Pesquisadores
- Amigos de Amigos
- Mais facilidade para realizar colaborações
- Cálculo da Distância dos nodos
- Algoritmo de Dijikstra
- os pesos representam força dos laços

Distância

Opsahl et al. 2010

$$d^{w\alpha}(i,j) = \min \left(\frac{1}{(w_{ih}^{\alpha})} + \dots + \frac{1}{(w_{hj}^{\alpha})} \right) \quad (3)$$

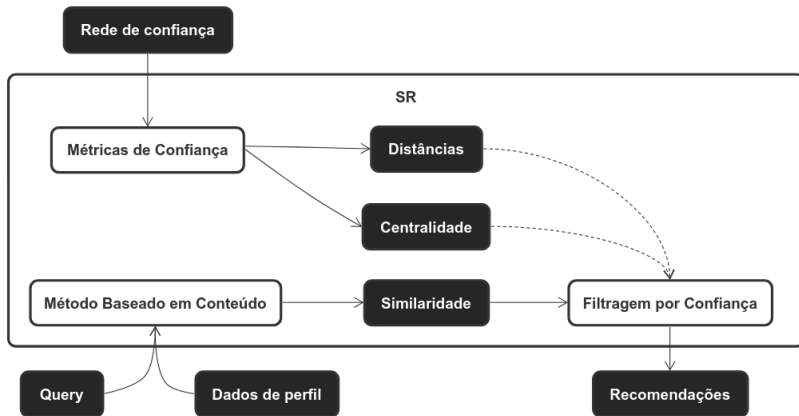
Caminho	$d^{w\alpha}(i,j)$			
	$\alpha = 0.00$	$\alpha = 0.50$	$\alpha = 1.00$	$\alpha = 1.50$
{A, B}	1.00	0.81	0.50	0.35
{A, C, B}	2.00	1.53	0.83	0.54
{A, D, E, B}	3.00	1.72	0.47	0.19

Pré-seleção de Recomendações

- Apenas confiança pode não ser suficiente para recomendações de qualidade
- Considerar também a linha de pesquisa do pesquisador
- Pré-seleção dos itens utilizando um método baseado em conteúdo
- Técnica de vetorização *TF-IDF*
 - Semelhança entre termos
- Sklearn

Arquitetura do SR

Massa & Avesani



Validação

Feedback dos usuários

- Estaria ele interessado em colaborar com o pesquisador recomendado ?
- Qual a facilidade de iniciar a colaboração ?
- Qual a confiança do usuário no pesquisador responsável ?

Conclusão

- É possível estender o trabalho com melhores heurísticas
 - Data de publicação
 - Eventos
- Incorporar mais métricas
 - Global
 - Local

TFG I - Proposta

Sistema de Recomendação Baseado em Confiança para
Promover a Colaboração em Redes de Pesquisa Científica

J.P. Saldanha¹ F. Prass¹

¹Ciência da Computação
Universidade Franciscana

Jul. 2019 / TFG I