

VRRC: Uma Ferramenta Web para Visualização e Recomendação em Redes de Coautoria

Eduardo M. Barbosa¹, Mirella M. Moro¹,
Giseli Rabello Lopes², J. Palazzo M. de Oliveira²

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)

²Instituto de Informática
Universidade Federal do Rio Grande do Sul (UFRGS)

{emb,mirella}@dcc.ufmg.br, {grlopes,palazzo}@inf.ufrgs.br

Resumo. *Este artigo propõe uma ferramenta Web que recebe de entrada qualquer lista de publicações e gera diversas visualizações e recomendações sobre a rede de coautoria formada a partir dessa lista. A ferramenta é uma solução prática e rápida para obter análises aprofundadas de uma rede de coautoria.*

Abstract. *This paper proposes a Web-based tool that receives any publication list and generates several visualizations and recommendations on the co-authorship network from this list. The tool is a practical and fast solution to obtain further analysis of a co-authorship network.*

1. Introdução

Geralmente, trabalhos científicos são realizados através de colaborações envolvendo pesquisadores com diferentes especialidades e perfis. Analisar tais colaborações e identificar novas possibilidades de parcerias não são tarefas triviais. Para a primeira, têm-se empregado análise de redes sociais para entender interconexões, evolução e comportamento [Ding 2011]. Exemplos variam de comunidades bem estabelecidas como a física até relativamente novas como a recuperação de informação. Métricas e técnicas de visualização em redes sociais vêm sendo aplicadas para o entendimento de um grupo mas, apenas recentemente, seu potencial para indicar novas colaborações começou a ser explorado. Em geral, é preciso usar mais de uma ferramenta para fazer análises e recomendações na rede e não há recursos para compartilhar os resultados obtidos de forma ágil.

Nesse contexto, propomos a ferramenta VRRC (pronuncia-se “ver-se”) para *Visualização e Recomendação em Redes de Coautoria*. A VRRC é uma solução Web prática para análise de qualquer rede de coautoria, e para recomendação de criação e intensificação de colaboração nessas redes. A ferramenta possibilita que qualquer usuário, a partir do fornecimento de dados sobre publicações, tenha acesso a um perfil completo na Web, com análises e recomendações visuais sobre a rede construída automaticamente com os dados fornecidos. Esse perfil pode ser compartilhado através de uma URL.

A recomendação automática gerada pela VRRC é útil em diversos cenários, por exemplo: um pesquisador que deseja identificar possibilidades de colaboração, e um jovem pesquisador que deseja engajar-se em uma nova comunidade. As visualizações da VRRC também podem ser aplicadas para análise da evolução temporal das colaborações.

Por exemplo, evolução das colaborações entre grupos compostos por membros de Programas de Pós Graduação ou pertencentes aos INCTs (Institutos Nacionais de Ciência e Tecnologia) [Laender et al. 2011]. No restante do artigo, apresentam-se: trabalhos relacionados (Seção 2), conceitos básicos (Seção 3), uma visão geral da ferramenta (Seção 4), as visualizações da rede e das recomendações (Seção 5) e as conclusões (Seção 6).

2. Trabalhos Relacionados

Existem vários grupos dedicados a aperfeiçoar sistemas de recomendação bem como expandir seus domínios para, por exemplo, recomendação de artigos [Hwang et al. 2010], mídias [Guy et al. 2010], entre outros. A pesquisa também tem explorado outros contextos, evoluindo da recomendação de itens para a recomendação de *indivíduos*. Um dos primeiros trabalhos nesse sentido introduz o *ReferralWeb* [Kautz et al. 1997], que identifica *especialistas* e define um caminho (em uma rede social) entre o usuário alvo e os indivíduos recomendados. De maneira similar, [McDonald 2003] detalha a avaliação de redes sociais que podem ser utilizadas para recomendar indivíduos.

As visualizações de redes, sem a recomendação, vêm sendo exploradas por sites como o *ArnetMiner* e o *Microsoft Academic Search*¹, que constroem a rede a partir de documentos minerados da Web e trabalham com um conjunto de dados pré-estabelecido, não permitindo que o usuário forneça seus próprios dados. Existem também ferramentas mais flexíveis, como o *Pajek* e o *Gephi*². Porém, elas não funcionam na Web, não permitem compartilhamento ágil das visualizações geradas e não tratam a geração da rede do usuário, a qual deve ser fornecida como entrada para o sistema.

Neste artigo, apresentamos a VRRRC, uma ferramenta para visualização e recomendação em redes de coautoria. As contribuições, em comparação ao estado-da-arte, podem ser resumidas da seguinte forma.

- A ferramenta VRRRC aceita entrada do usuário; centraliza visualizações da rede e de recomendação em um mesmo perfil; é uma ferramenta de acesso via Web; e permite fácil compartilhamento das visualizações geradas. As ferramentas existentes não combinam todas essas funcionalidades.
- Enquanto as abordagens tradicionais apresentam os resultados da recomendação através de listas ou tabelas, a VRRRC define representações gráficas para facilitar a visualização de tais resultados.
- As abordagens atuais de recomendação apenas indicam novas colaborações. Com a VRRRC, o usuário pode identificar não apenas novas colaborações a serem definidas como também colaborações existentes que podem ser intensificadas.

3. Conceitos

Esta seção apresenta os principais conceitos envolvidos na definição das visualizações e das recomendações geradas pela ferramenta VRRRC.

Uma **rede de coautoria** pode ser caracterizada por um grafo onde um nodo n_i [$0 < i \leq Num_autores$] representa um autor, e uma aresta $A(n_{i_a}, n_{i_b}, p)$ [$i_a \neq i_b$] a colaboração entre os autores a e b , com o peso p . Se dois autores nunca publicaram

¹ArnetMiner: www.arnetminer.com; Microsoft Academic Search: academic.research.microsoft.com

²Pajek: vlado.fmf.uni-lj.si/pub/networks/pajek; Gephi: gephi.org

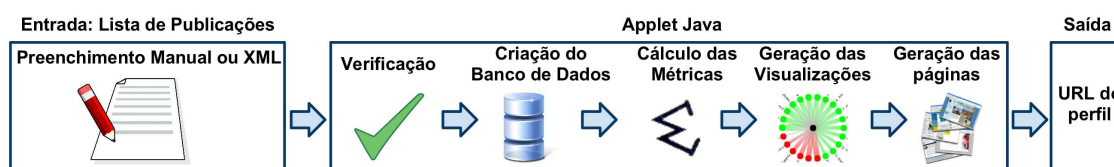


Figura 1. Visão geral da ferramenta VRRC

juntos, não há aresta conectando-os. O peso da aresta indica a intensidade da colaboração entre os autores. Como grafo não-direcional, o peso p pode ser o número de publicações conjuntas. Como grafo bidirecional, p é a métrica de *Cooperação* apresentada a seguir.

Duas métricas definidas por [Lopes et al. 2010] são utilizadas pela VRRC para determinar recomendações no contexto acadêmico. A métrica de **Cooperação** quantifica a cooperação existente com base nos artigos em coautoria, e a métrica de **Correlação** quantifica a correlação entre os pesquisadores considerando suas diferentes áreas de atuação. Em resumo, a *Cooperação* do autor a com b é um valor entre 0 e 1 definido pela divisão do número de artigos em colaboração entre a e b pelo total de artigos de a (note que o valor é diferente para a cooperação de b com a); e a *Correlação* é definida por uma equação mais complexa, que leva em conta o perfil de publicação (as áreas nas quais os autores publicam) e utiliza o modelo de espaço vetorial (*VSM*) para calcular os valores globais entre pares de autores. A relação entre *Cooperação* e *Correlação* entre os pares estabelece a necessidade (ou não) de uma maior interação de pesquisa entre esses. Uma recomendação é gerada quando existir média ou alta *Correlação* para baixa *Cooperação*.

4. Visão Geral

A ferramenta VRRC possui uma interface Web e permite que qualquer pessoa crie um perfil para obter as visualizações da rede desejada. É necessário apenas fornecer os dados sobre publicações que servirão de base para geração da rede de coautoria. Tais dados podem ser informados: manualmente através do preenchimento de uma lista de publicações; ou automaticamente através de um arquivo XML com metadados descritos no formato *Dublin Core* (DC)³. As informações mínimas necessárias sobre cada publicação e os elementos do DC correspondentes são: título (*title*), ano (*date*), área de pesquisa (*subject*) e autores (*creator*). Alternativamente, a ferramenta também permite que bases de dados pré-carregadas possam ser compartilhadas⁴.

A Figura 1 mostra uma visão geral do funcionamento da VRRC. A criação do perfil do usuário é feita por um *Applet JAVA* que recebe uma lista de publicações, verifica se todos os dados necessários estão nessa lista, e cria e popula um banco de dados MySQL para o perfil. O banco de dados é constituído de três entidades principais: *Autor*, *Área de Pesquisa* e *Publicação*. *Área de Pesquisa* possui uma relação *1:n* com *Publicação*, que por sua vez, possui uma relação *n:m* com *Autor*. Esse esquema foi definido para possibilitar a aplicação automática das métricas descritas na Seção 3. Após, a ferramenta calcula essas métricas e gera as visualizações e páginas Web para navegação pelo perfil. Como resultado, é retornada uma URL de acesso ao perfil. Outros detalhes, sobre as visualizações geradas pela ferramenta, são apresentados na Seção 5.

³XML: <http://www.w3.org/XML/>; DC: <http://dublincore.org/>; Esquema XML do DC: <http://dublincore.org/schemas/xmls/simpledc20021212.xsd>

⁴Por exemplo, a base do *Ciência Brasil* [Laender et al. 2011] e da DBLP: <http://dblp.uni-trier.de/>

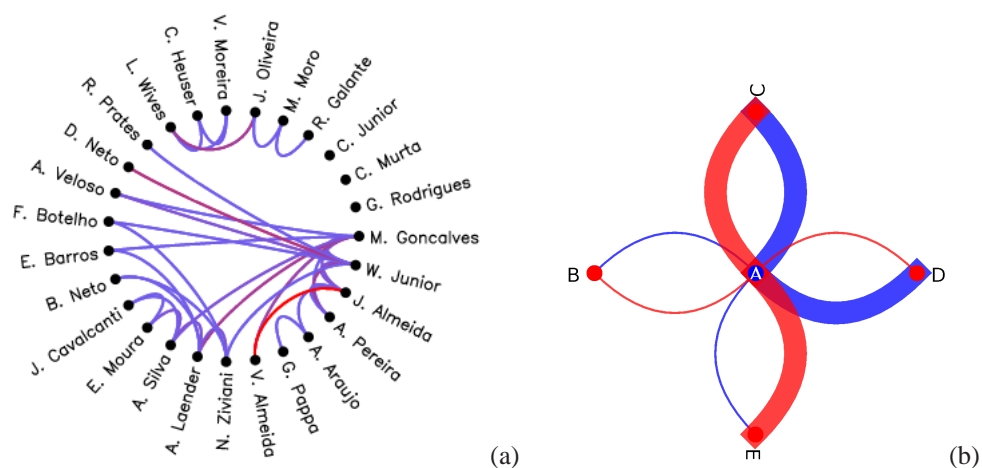


Figura 2. (a) Visualização global da rede do InWeb em 2008; e (b) Visualização da rede centrada no autor A

5. Visualizações da VRRC

Para demonstrar as funcionalidades da VRRC, são apresentadas as principais visualizações disponibilizadas pela ferramenta: relacionadas à representação da rede de coautoria (Seção 5.1) e à apresentação das recomendações geradas (Seção 5.2).

5.1. Visualização da rede de coautoria

Visualização global da rede. A visualização global da rede objetiva fornecer uma representação completa da rede de coautoria. Os autores são distribuídos uniformemente em uma circunferência, e a colaboração é representada por uma aresta colorida, indicando a sua intensidade. A cor da aresta pode variar de azul (colaboração mais fraca) até vermelho (colaboração mais forte). As arestas de um nodo são destacadas ao passar o mouse sobre ele. A Figura 2a mostra essa visualização para a rede do InWeb⁵ formada com dados de publicações de 2008 coletados dos Currículos Lattes dos pesquisadores. A implementação dessa visualização foi feita a partir de uma adaptação da *MooWheel*⁶ (Javascript). Essa adaptação posiciona os nodos de forma a evidenciar clusters de pesquisadores, quando existentes, e permite filtragem das arestas por intervalo de tempo (em anos), conforme solicitação do usuário.

Visualização da evolução da rede. Esta visualização fornece uma visão geral da evolução das colaborações de uma rede, considerando um comparativo entre diferentes intervalos de tempo. Os autores são distribuídos uniformemente em uma circunferência, e a colaboração é representada por uma aresta colorida (semelhante à Figura 2a), indicando uma das seguintes situações: arestas *verdes* para novas colaborações surgidas; *vermelhas* para colaborações que foram intensificadas; e *pretas* para colaborações que não sofreram alterações no período especificado.

Visualização da rede centrada no autor. Esta visualização objetiva exibir a importância entre autores definida pela métrica de *Cooperação* (Seção 3). Essa exibição, considerando toda a rede, pode ser ilegível, devido ao provável grande número de autores, bem como de

⁵Instituto Nacional de Ciência e Tecnologia para a Web: <http://www.inweb.org.br>

⁶MooWheel: <http://labs.unwielydy.net/moowheel>

ligações entre eles. Por isso, a ferramenta VRRRC permite a definição de várias subredes, cada uma contendo apenas um autor principal e seus coautores. O autor principal é posicionado no centro do grafo e seus coautores são distribuídos uniformemente ao redor. Entre cada par de autores existem dois arcos, um para representar a importância do autor central para seu coautor e outro para a relação contrária. Usando uma analogia a um relógio, o arco curvado para a posição de menor hora (vermelho) representa a importância do coautor para o autor central, e o outro arco (azul) representa a relação contrária. A intensidade da importância é representada pela espessura do arco. A Figura 2b contém um exemplo dessa visualização, sendo possível inferir diversas interpretações sobre as relações.

- *A e B*: *A* tem pouca importância para *B* e vice-versa, ou seja, *A* e *B* escreveram poucos artigos juntos em relação à quantidade de artigos que escreveram com outros pesquisadores. Essa é uma relação típica, por exemplo, entre dois professores de universidades distintas que têm pouca oportunidade de colaborar entre si.
- *A e C*: *A* tem muita importância para *C* e vice-versa. Provavelmente, esses pesquisadores publicaram aproximadamente a mesma quantidade de artigos e compartilham coautoria em uma parte considerável desses.
- *A e D*: relação de extremos opostos, onde *A* tem muita importância para *D*, mas *D* tem pouca importância para *A*. A quantidade de artigos que *A* escreve com *D* é pouco significativa tendo em vista todas as publicações de *A*, entretanto, quase todas as publicações de *D* são conjuntas com *A*. Essa relação pode caracterizar, por exemplo, a de um orientador experiente *A* e um de seus jovens orientandos *D*.
- *A e E*: situação simetricamente análoga à relação entre *A* e *D*.

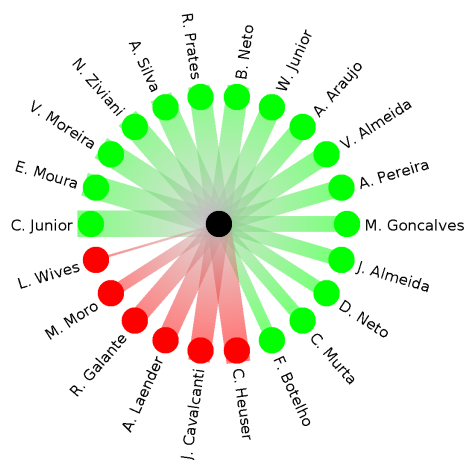
5.2. Visualização de recomendação

As visualizações, descritas nesta seção, objetivam apresentar as recomendações geradas pela VRRRC, utilizando as métricas de *Cooperação* e *Correlação*.

Visualização de recomendação centrada no autor. Esta visualização exibe as recomendações de criação e intensificação de colaboração para um autor específico. Esse autor é situado no centro do grafo e os autores relacionados são distribuídos ao seu redor. A espessura da aresta indica a intensidade da recomendação, de forma que uma aresta mais espessa indica uma recomendação mais expressiva. A cor de apresentação de cada nodo (e da aresta que liga este ao nodo central) pode ser uma das seguintes: *verde* para recomendação de início de colaboração; *vermelho* para recomendação de intensificação de colaboração já existente; e *preto* para aquelas colaborações que já estão ocorrendo em intensidades adequadas ou superiores aos graus de correlação encontrados (ou seja, não há necessidade de recomendação). A Figura 3 apresenta um exemplo dessa visualização para o pesquisador *J. Oliveira*, na rede do InWeb, com dados de publicações obtidos até 2010, coletados de currículos Lattes em 29 de novembro de 2010.

Visualização global de recomendação. A visualização global de recomendação tem como objetivo exibir as recomendações para todos os pesquisadores da rede. A forma de apresentação dessa visualização é a mesma da Figura 2a. A diferença está no significado das arestas, que podem ser apenas *verdes* ou *vermelhas* para representar, respectivamente, recomendação de início de colaboração ou intensificação de colaboração já existente. Nesse caso, são representadas apenas as ligações em que ocorre recomendação a pelo menos um dos dois autores envolvidos na relação.

Pesquisador Principal: J. Oliveira



As recomendações apresentadas estão ranqueadas por ordem de importância, sendo que a primeira recomendação de nova colaboração é a do pesquisador *C. Junior* (aresta verde mais espessa), e a primeira recomendação de intensificação de colaboração é a do pesquisador *C. Heuser* (aresta vermelha mais espessa). Nesse exemplo, não há nenhuma aresta preta, indicando que nenhuma das colaborações já existentes tem intensidade suficiente, por isso, foram marcadas de vermelho e devem ser intensificadas.

Figura 3. Visualização de recomendação centrada no autor *J. Oliveira*

6. Conclusão

A VRRC, ferramenta desenvolvida neste trabalho, é muito prática e provê uma ótima estrutura para compreensão e análise da rede de coautoria de qualquer grupo de pesquisa. Além disso, permite fácil compartilhamento das visualizações geradas, pois opera em plataforma Web. Espera-se que a VRRC seja amplamente usada, principalmente para aplicações na área acadêmica. O uso da ferramenta para análises de grupos de pesquisa pode fornecer indícios importantes para orientar o crescimento de interações entre pesquisadores, de forma que se alcancem novos projetos e publicações conjuntas.

Agradecimentos. Trabalho financiado pelos projetos InWeb (CNPq 573871/2008-6) e Chasque (CNPq 483397/2010-5), e pelas bolsas individuais dos autores.

Referências

- Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of co-authorship and citation networks. *Journal of Informetrics*, 5(1):187 – 203.
- Guy, I. et al. (2010). Social media recommendation based on people and tags. In *Procs. of ACM SIGIR*, pages 194–201, Geneva, Switzerland.
- Hwang, S.-Y. et al. (2010). Coauthorship networks and academic literature recommendation. *Electron. Commer. Rec. Appl.*, 9(4):323–334.
- Kautz, H. et al. (1997). Referral Web: combining social networks and collaborative filtering. *Commun. ACM*, 40(3):63–65.
- Laender, A. H. F. et al. (2011). Building a Research Social Network from an Individual Perspective. In *Procs. of JCDL, Poster*, Ottawa, Canada.
- Lopes, G. R. et al. (2010). Collaboration Recommendation on Academic Social Networks. In *Procs. of ER 2010 Workshops*, pages 190–199, Vancouver, Canada.
- McDonald, D. W. (2003). Recommending collaboration with social networks: a comparative evaluation. In *Procs. of SIGCHI*, pages 593–600, Fort Lauderdale, USA.