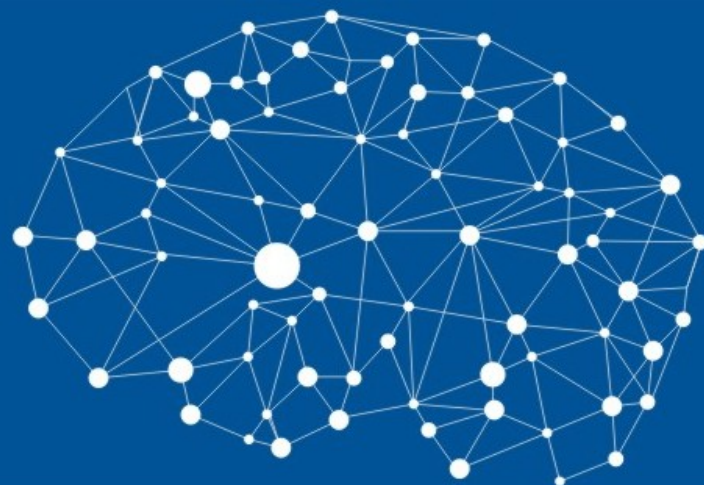




Uma estratégia baseada em árvores genealógicas científicas para visualização da relação orientador-orientado

Tales Henrique José Moreira¹; Thiago Magela Rodrigues Dias²; Gray Farias Moita³

MOREIRA, T. H. J.; DIAS, T. M. R.; MOITA, G. F.. Uma estratégia baseada em árvores genealógicas científicas para visualização da relação orientador-orientado In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. **Anais...** São Paulo: USP, 2016. p. A91

^{1, 2, 3} Centro Federal de Educação Tecnológica de Minas Gerais

UMA ESTRATÉGIA BASEADA EM ÁRVORES GENEALÓGICAS CIENTÍFICAS PARA VISUALIZAÇÃO DA RELAÇÃO ORIENTADOR-ORIENTADO

Eixo temático: Colaboração na Ciência

Modalidade: Apresentação oral

1 INTRODUÇÃO

A evolução das pesquisas científicas têm forte influência do processo de formação, onde pesquisadores orientadores inserem novos pesquisadores que contribuem para que novos estudos sejam realizados em diversas áreas do conhecimento. Grande parte dos trabalhos realizados com orientação no Brasil são decorrentes de Programas de Pós-Graduação (PPGs), impulsionados pela necessidade de capacitação e titulação de docentes e pesquisadores.

Para Ferreira et al. (2009), o binômio ou díade orientador-orientado é indubitavelmente a base dos PPGs, o que determina o crescimento e a expansão dos cursos de Pós-Graduação (PG) e a demanda de orientação. Além disso, os autores ressaltam que o aluno de PG é um pesquisador em potencial, em estágio avançado de desenvolvimento, ou seja, a caminho da autonomia científica, mas ainda dependente de um professor, o que justifica as atividades de orientação como efetivamente necessárias.

Para Sugimoto (2014), genealogia acadêmica é o estudo quantitativo da herança intelectual adquirido através do relacionamento orientador-orientado. Logo, compreender tal relação pode proporcionar mecanismos para analisar como as orientações estão sendo conduzidas. Diante disso, este trabalho propõe uma abordagem baseada em árvores para a modelagem da genealogia científica de pesquisadores orientadores, possibilitando visualizar todo o histórico de orientação de forma hierárquica, e consequentemente como o seu conhecimento tem se propagado. Com isso, diversos outros trabalhos podem ser propostos objetivando analisar as árvores genealógicas científicas caracterizadas.

Árvores genealógicas científicas são caracterizadas como árvores que representam hierarquicamente o histórico de um orientador e todos os seus orientados. Logo, caracterizando uma rede genealógica científica é possível observar todo o histórico de um

determinado orientador e como o seu conhecimento foi repassado ao longo do tempo (MOREIRA et al., 2014).

Adicionalmente aos dados básicos que caracterizam os vínculos, informações como quantidade de publicações, área de atuação e projetos de pesquisa podem ser extraídos de todos os elementos que compõem a rede a ser analisada com o intuito de se obter dados quantitativos que representam a difusão do conhecimento a partir de um orientador em particular.

Para isso, dados sobre orientações bem como de características individuais de cada um dos indivíduos que compõem a rede são analisados. Neste trabalho, os dados utilizados são os currículos que compõem a Plataforma Lattes sob a coordenação do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Os currículos que compõem a plataforma são analisados a partir de um determinado orientador que representa a raiz da árvore. Este processo é realizado de forma recursiva, com o objetivo de gerar uma árvore integrada com dados de produção bibliográfica, permitindo a visualização da mesma e intensidade de produção em coautoria.

2 TRABALHOS RELACIONADOS

Em Filho e Martins (2006), são verificadas as influências da relação orientador-orientado no processo de produção de teses e dissertações dos PPGs em Contabilidade da cidade de São Paulo. Os autores citam como justificativas para o estudo a importância de se analisar aspectos que teriam ligação com a construção do conhecimento, especificamente em se tratando da área de Contabilidade e a tentativa de sinalizar a importância da temática orientação.

Tuesta et al. (2012) apresentam uma análise temporal da relação orientador-orientado, com um estudo de caso sobre a produtividade dos pesquisadores doutores da área de Ciência da Computação extraindo os dados de análise dos currículos Lattes. No trabalho, são analisadas as principais características do grupo e as relações de coautoria.

Diversos outros trabalhos têm explorado a Plataforma Lattes como principal fonte de informações para caracterização e análise de orientações (Miyahara, 2011; Mena-Chalco e Cesar-Junior, 2011; MOREIRA et al., 2014).

Diferentemente dos trabalhos citados acima, este visa gerar árvores genealógicas científicas através de um algoritmo recursivo e principalmente, vincular informações a cerca do relacionamento orientador-orientado, como áreas de atuação, produção bibliográfica realizada em colaboração, entre outras.

3 METODOLOGIA

Este trabalho tem como principal fonte de dados os currículos da Plataforma Lattes. A Plataforma Lattes representa a experiência do CNPq na integração de bases de dados de Currículos, de Grupos de pesquisa e de Instituições em um único Sistema de Informações, se tornando um padrão nacional no registro da vida pregressa e atual dos estudantes e pesquisadores do país (CNPQ, 2015).

Apesar de os dados dos currículos da Plataforma Lattes serem disponibilizados livremente, esses são visualizados através de interface de consulta que apresenta os currículos individualmente. Diante disso, técnicas e ferramentas para a extração e integração dos dados com outras bases de dados científicos para complementação das informações se fazem necessárias (MOREIRA et al., 2014). Para a extração dos currículos, foi utilizado o framework de extração e integração de dados científicos, desenvolvido por Dias et al. (2013) para obtenção dos dados a serem analisados.

Os currículos são extraídos pelo framework e armazenados localmente em XML (Extensible Markup Language) para futuras análises. Para a construção das árvores, uma estratégia recursiva é utilizada, na qual uma raiz é previamente selecionada, fazendo a identificação de todas as orientações e supervisões concluídas, obtendo uma lista de orientados que estão diretamente vinculados ao nó raiz. Este processo é repetido para cada item da nova lista, identificando assim os vários níveis de descendentes do nó raiz. Consequentemente, currículos que não possuem orientações concluídas vão se tornando nós folhas da árvore. Esse processo acontece até que todos os descendentes sejam processados, encontrando por fim apenas nós folhas.

De posse desses dados, foram obtidos os trabalhos realizados em conjunto, a partir da relação orientador-orientado. Para a identificação da produção do orientador proveniente de colaboração com o aluno orientado, foi utilizado o módulo de Identificação e Caracterização

também proposto pelo framework (DIAS et al., 2013). Utilizando o referido módulo, os dados foram processados, produzindo como resultado a rede de colaboração científica do orientador. Tais resultados consistem em nós e arestas, que possibilitam a análise da rede caracterizada com técnicas de análise de redes sociais, possibilitando obter conhecimento sobre como acontece a colaboração científica dos pesquisadores analisados.

4 RESULTADOS

A partir dos dados processados, é possível obter informações como o identificador do currículo, o nome do pesquisador, sua afiliação atual, o nível da orientação que foi utilizada na identificação, além de outras informações que podem ser extraídas dos currículos com o intuito de fornecer um conjunto maior de informações que poderão ser analisadas. Observa-se que nem todos os nós possuem identificador, isso ocorre quando o registro da orientação presente no currículo do orientador não possui vínculo do nome do orientado e seu ID Lattes, essa situação pode ser visualizada na Figura 1. Com isso, nós sem vínculo (ID Lattes) são inseridos apenas com o nome como nós folhas, ou seja, terminais, tendo em vista que sem o ID Lattes não é possível acessar o currículo individual de um orientador e consequentemente não é possível analisar as suas orientações.

Figura 1 - Trecho do currículo do orientador exibindo orientados vinculados pelo identificador e orientados não vinculados

- Tese de doutorado**
- Não vinculado**
1. Aline de Sousa Pereira Bitencourt. Desenvolvimento de métodos automatizados para estimação de parâmetros de algoritmos heurísticos. Início: 2014. Tese (Doutorado em Modelagem Matemática e Computacional) - Centro Federal de Educação Tecnológica de Minas Gerais. (Orientador).
- Vinculado**
2.  Thiago Magela Rodrigues Dias. Segurança da Informação: Descoberta de Conhecimento em Fontes de Dados. Início: 2013. Tese (Doutorado em Modelagem Matemática e Computacional) - Centro Federal de Educação Tecnológica de Minas Gerais, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. (Orientador).
- Vinculado**
3.  Henrique Costa Braga. Simulação da movimentação de pessoas em situações de emergência: uma abordagem topológica com autômatos celulares fuzzy. Início: 2013. Tese (Doutorado em Modelagem Matemática e Computacional) - Centro Federal de Educação Tecnológica de Minas Gerais, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. (Orientador).

Fonte: Fragmento currículo Lattes

Com o processamento dos dados das árvores geradas e das arestas obtidas com o módulo de identificação do Framework de Extração, podem ser gerados relatórios com a produção científica realizada em colaboração do orientador-orientado, para futuras análises.

Analisando a rede do pesquisador e orientador Gray Farias Moita, utilizado como raiz no exemplo deste trabalho, descobriu-se que 30,18% da produção do orientador é proveniente de seus orientados, aumentando consideravelmente os números de publicações do orientador. Para outros vinte e três orientados deste orientador não foram encontradas publicações, ou seja, mesmo tendo orientado alunos, estes não publicaram em conjunto.

Tabela 1 – Número de publicações obtidas através das cinco orientações mais produtivas do pesquisador Gray

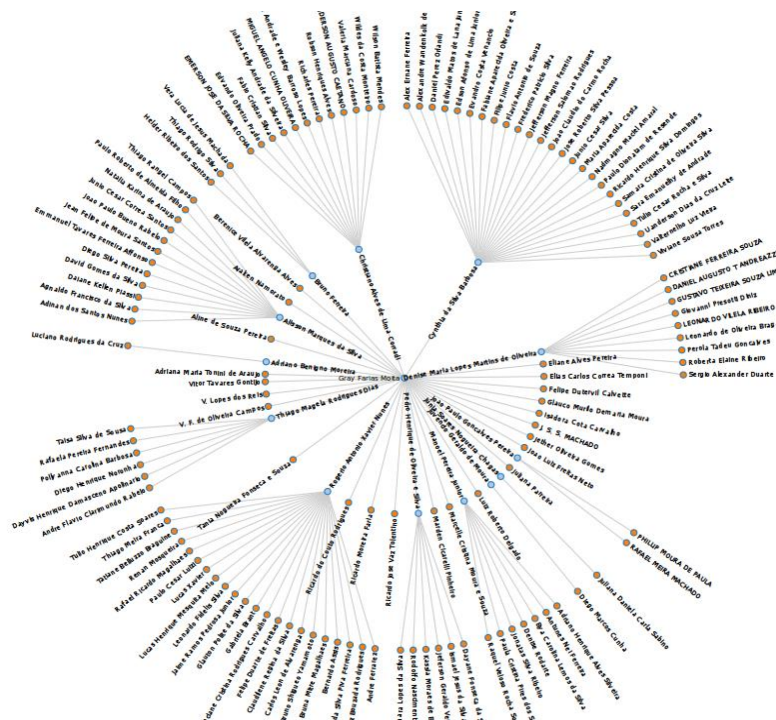
Orientados:	Produção em Colaboração
Thiago Magela Rodrigues Dias	49
Junia Soares Nogueira Chagas	7
Bruno Ferreira	6
Juvencio Geraldo de Moura	5

Fonte: Próprios Autores

Além da tabela de orientações e dos dados de produção científica, as árvores genealógicas podem ser modeladas em diversos formatos, possibilitando a análise em diversas ferramentas de visualização.

Na figura 2, é possível observar como estão estruturados hierarquicamente todos os descendentes de um orientador que é o centro da árvore gerada. Cada nível da árvore representa um salto na distância entre os nós.

Figura 2 - Árvore genealógica acadêmica estática



Fonte: Próprios Autores

5 CONCLUSÕES

A construção de árvores genealógicas científicas é uma técnica interessante para análise visual de como ocorre a propagação do conhecimento de um orientador. A representação fornece de forma hierárquica e estruturada todo o histórico de orientações, facilitando a visualização e análise da relação orientador-orientado.

Neste trabalho, foram utilizados métodos que permitem gerar árvores genealógicas através de todos os pesquisadores da base de dados de forma recursiva e integrado com outras ferramentas já desenvolvidas. Além disso, métodos de visualização das árvores criadas também foram utilizados. Isso demonstra flexibilidade e aplicabilidade do método, possibilitando gerar e estruturar os dados de acordo com a necessidade, ou da ferramenta utilizada para análise e visualização da árvore.

Conforme apresentado neste trabalho, outra grande aplicação das árvores genealógicas é identificar o ganho de produção dos pesquisadores orientadores através de trabalhos publicados com seus orientados.

AGRADECIMENTOS

Os autores agradecem o apoio da FAPEMIG, bem como ao CEFET-MG, para a realização desse trabalho.

REFERÊNCIAS

CNPq, <http://www.cnpq.br/pt/web/portal-lattes/sobre-a-plataforma>. Acesso em 08/03/2015.

DIAS, T. M. R. et al. Modelagem e Caracterização de Redes Científicas: Um Estudo Sobre a Plataforma Lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, II, Maceió, 2013.

Dias, T. M. R.; Moita, G. F. Extração e Modelagem de Redes de Colaboração Científica. In CONFERÊNCIA IADIS IBERO-AMERICANA WWW/INTERNET, Porto Alegre, Brasil, 2013.

FERREIRA, L. M.; FURTADO, F.; SILVEIRA, T. S. Relação orientador-orientando: o conhecimento multiplicador. *Acta Cir.* Brasil., São Paulo, v. 24, n. 3, 2009.

LEITE-FILHO, G. A.; MARTINS, G. A. Relação orientador-orientado e suas influências na elaboração de teses e dissertações. *RAE*. Edição especial. 2006.

MENA-CHALCO, J. P., CESAR-JUNIOR, R. M. (2011). Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. Capítulo do livro **Bibliometria e Cientometria: reflexões teóricas e interfaces** (in press). São Carlos: Pedro & João, 1-20.

MIYAHARA, E. K., MENA-CHALCO, J. P.; CESAR-JUNIOR, R. M. 2011. Genealogia Acadêmica Lattes. Monografia. Universidade de São Paulo. Instituto de Matemática e Estatística. São Paulo.

MOREIRA, T. H. J.; DIAS, THIAGO. M. R. ; SANTOS, L. R. F. ; MOITA, G. F. . Genealogia Científica: Uma Análise Hierárquica de pesquisadores orientadores. In: XI Simpósio de Mecânica Computacional (SIMMEC), Juiz de Fora, 2014.

SUGIMOTO, C. R. (Eds.). **Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact**. Cambridge, MA: The MIT Press, 2014. p. 365–382.

TUESTA, E. F. et al. Análise temporal da relação orientador-orientado: um estudo de caso sobre a produtividade dos pesquisadores doutores da área de ciência da computação. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, II, Curitiba, 2012.