

# Um estudo da colaboração científica brasileira com adoção de técnicas de análises de redes

Thiago Magela Rodrigues Dias<sup>1</sup>; Gray Farias Moita<sup>2</sup>; Patrícia Mascarenhas Dias<sup>3</sup>

## Fundamentações

Nos últimos anos, além da análise bibliométrica de publicações científicas, diversos outros estudos têm procurado compreender como a ciência tem evoluído e como a colaboração científica ocorre. Para isso, diversas técnicas baseadas em análises de redes surgem como uma alternativa para verificar esse fenômeno.

Newman (2001) apresenta uma avaliação extensa sobre características sociais das redes de colaboração científica nas áreas de computação, biologia, física e medicina no período de 1995 a 1999, encontrando diversas características sociais que são comuns a essas redes. Já em outro trabalho (Newman, 2004), o autor analisa redes de coautoria nas áreas de biologia, física e matemática visando

---

1 Doutor, Centro Federal de Educação Tecnológica de Minas Gerais – CEFET-MG, Brasil, thiagomagela@cefetmg.br.

2 Doutor, Centro Federal de Educação Tecnológica de Minas Gerais – CEFET-MG, Brasil, gray@dppg.cefetmg.br.

3 Mestre, Centro Federal de Educação Tecnológica de Minas Gerais – CEFET-MG, Brasil, patriciamdias@gmail.com.

identificar propriedades estatísticas, como número médio de publicações e quantidade média de autores, de modo a determinar padrões comuns entre os elementos da rede analisada.

Com a modelagem e caracterização das redes de colaboração científica, é possível aplicar diversas técnicas que permitem entender como essas redes são estruturadas, fornecendo assim subsídios para diversos estudos como predição de vínculos entre pesquisadores, recomendação de especialistas e identificação de grupos de pesquisa. Porém, a identificação dessas colaborações em grandes repositórios de dados não é uma tarefa trivial devido a uma série de fatores, como ambiguidade entre nomes de autores, relação de autores incompleta, erros ortográficos nas citações ou nos títulos das publicações, dentre outros.

## **Objetivos**

Este trabalho tem como objetivo apresentar um estudo sobre as redes de colaboração científica dos doutores brasileiros, baseado em análises de redes que possibilitam compreender como ocorre a colaboração científica do conjunto em análise e como estão colaborando os pesquisadores nas diversas áreas do conhecimento. Para isso, são analisados dados dos currículos cadastrados na Plataforma Lattes, responsável por armazenar informações curriculares da comunidade científica nacional. A Plataforma Lattes é considerada uma importante fonte de dados sobre a produção científica brasileira.

## Método

Considerando que os trabalhos correlatos analisaram em sua maioria apenas grupos específicos de indivíduos, tendo em vista que a manipulação de grandes quantidades de currículos da Plataforma Lattes não é uma tarefa trivial, já que existem problemas que envolvem recuperação de informação e algoritmos eficientes para manipulação de grande volume de dados, a ferramenta *LattesDataXplorer* (Dias, 2016), um *framework* para extração e tratamento dos dados foi utilizado (Figura 1).

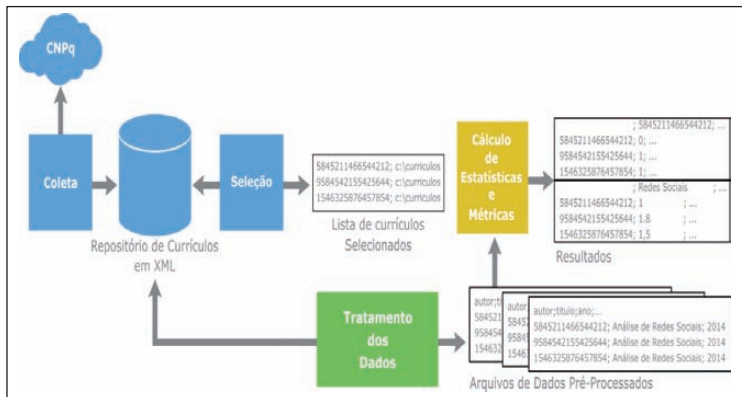


Figura 1. *LattesDataXplorer*, framework para extração e tratamento de dados curriculares da Plataforma Lattes (Dias, 2016).

O processo de coleta de todos os dados curriculares da Plataforma Lattes é dividido em três componentes que objetivam minimizar o custo computacional: 1) extração de URLs, que é responsável por extrair as referências únicas para todos os currículos cadastrados e desta forma possibilitar o acesso individual a cada currículo, 2) extração

de Identificadores (Ids) e Data, que visa acessar cada currículo e extrair o seu identificador individual, bem como a data de última atualização, 3) extração de currículos, que é responsável por extrair e armazenar os currículos (em formato XML) cuja data de atualização na Plataforma Lattes seja divergente da data de atualização do currículo armazenado localmente.

Todas essas etapas se fazem necessárias, já que o ideal é manter os dados curriculares atualizados com a maior frequência possível, possibilitando a realização de análises com dados atualizados, tendo em vista que, com a estratégia adotada, não se faz necessário extrair todo o repositório de dados a cada nova extração. Além dos componentes responsáveis pelo processo de coleta dos dados, e considerando a necessidade de análises de grupos específicos, como por exemplo, de todos os indivíduos com doutorado concluído, um componente denominado Seleção, caracteriza-se como importante mecanismo no contexto deste trabalho.

O componente de Seleção utiliza a linguagem de consulta XPath (XML Path Language) para pesquisa no repositório local de currículos e posterior geração do conjunto. A linguagem XPath permite construir expressões que vão processar e percorrer um documento XML de forma similar ao uso de expressões regulares. Logo, é possível agrupar um conjunto de currículos com base em parâmetros desejados.

Com auxílio da ferramenta foi possível identificar um conjunto de 223.477 indivíduos que possuem o nível de formação doutorado concluído, registrado em seus currículos. Logo, os dados contidos nos currículos deste conjunto de

indivíduos foram analisados, sendo apresentada uma visão geral deste conjunto, e posteriormente, são caracterizadas as redes de colaboração científica para análises.

## Resultados

Neste trabalho, são analisadas as redes de colaboração científica dos doutores brasileiros, considerando suas grandes áreas de atuação no período de 1960 a 2014, período este que compreende o ano de publicação da produção mais antiga cadastrada na Plataforma Lattes ao último ano do período de análise considerado. Foi possível observar que uma grande quantidade (35,85%) dos artigos publicados em anais de congresso e em periódico são de autoria única e, não caracterizam colaborações.

A análise da colaboração científica, entretanto, pode ser mais bem evidenciada com a caracterização das respectivas redes de colaboração. Neste trabalho, foram caracterizadas redes temporais e, para suas análises, algumas métricas de análise de redes sociais foram aplicadas.

Para a identificação das colaborações, um método denominado ISColl, foi utilizado (Dias e Moita, 2015). Para a caracterização das redes que contêm todos os elementos, foram adotados métodos que visam concentrar os nós mais conectados no centro da rede e, conseqüentemente, aqueles nós com menor quantidade de ligações ou isolados são deslocados para as extremidades das redes (Figura 2).

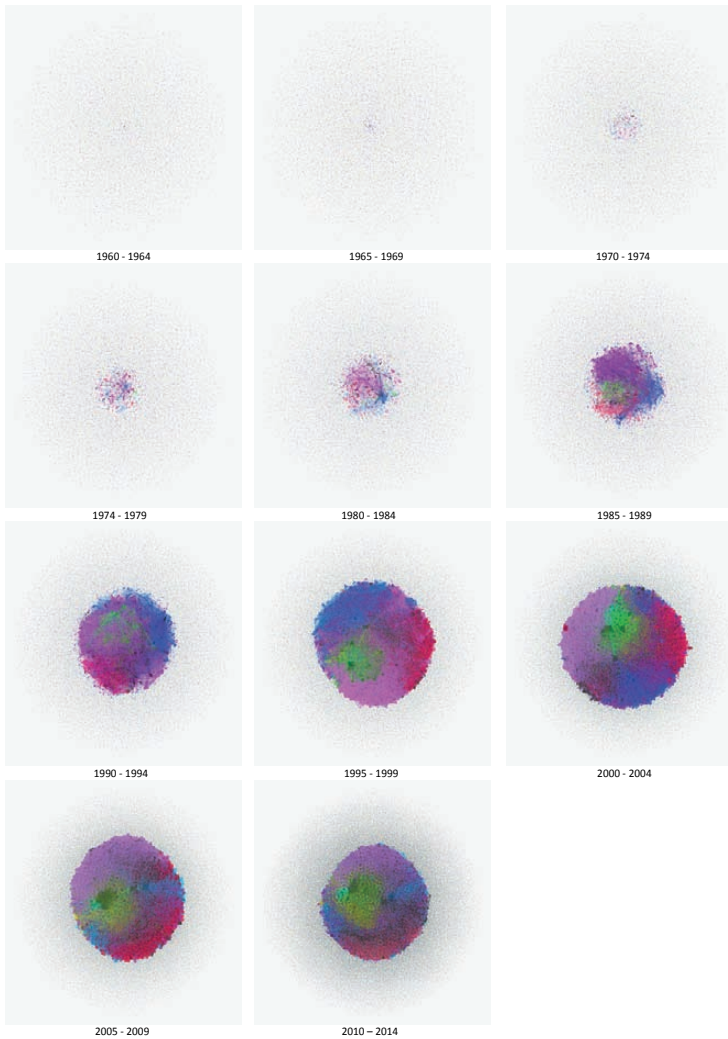


Figura 2. Redes gerais de colaboração científica contendo todos os nós.

As colaborações entre os doutores se iniciam no início da década de 1960 e são intensificadas a partir da década de 1990. É possível observar como as colaborações em al-

gumas grandes áreas se estabilizaram ao longo do tempo, por exemplo, nas grandes áreas de Ciências da Saúde e Ciências Biológicas, bem como outras que passaram a ter um maior grau de colaboração nos últimos anos, como por exemplo, as grandes áreas de Ciências Exatas e da Terra e Ciências Agrárias.

## Conclusões

Foi possível observar como os pesquisadores de determinadas grandes áreas do conhecimento tendem a realizar seus trabalhos de forma mais colaborativa e, ainda, como estas grandes áreas intensificaram seus trabalhos em colaboração em cada período analisado. Além disso, verificando as arestas mais densas das redes caracterizadas, foi identificado como pequenos grupos têm trabalhado em colaboração de forma intensa, aumentando de forma impactante o número de publicações de seus pesquisadores.

## Referências

- Dias, T. M. R., e Moita, G. F. A method for the identification of collaboration in large scientific databases. *EmQuestão*, v. 21, p. 140–161, 2015.
- Dias, T. M. R. *Um Estudo Sobre a Produção Científica Brasileira a partir de dados da Plataforma Lattes*. 2016. 181 f. Tese (Doutorado) – Curso de Programa de Pós-graduação, Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2016.
- Newman, M. E. J. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, v. 98, n. 2, p. 404–409, 2001.

Newman, M. E. J. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, v. 101, suppl 1, p. 5200–5205, 2004..