

## IDENTIFICATION SHORTEST PATHS IN SCIENTIFIC NETWORKS

Felipe Affonso (Instituto de Ensino Superior e Pesquisa – INESP, MG, Brasil) –  
felipe-affonso@hotmail.com

Thiago Magela Rodrigues Dias (Centro federal de Educação Tecnológica de  
Minas Gerais – CEFET-MG, MG, Brasil) – thiagomagela@gmail.com

Patrícia Mascarenhas Dias (Instituto de Ensino Superior e Pesquisa – INESP,  
MG, Brasil) – pmdias@gmail.com

Gray Farias Moita Centro federal de Educação Tecnológica de Minas Gerais –  
CEFET-MG, MG, Brasil) – gray@dppg.cefetmg.br

**Abstract.** *Scientific social networks have attracted attention of researchers seeking to understand their evolution and behavior. The amount of data in these networks makes them valuable source of information. This paper seeks to identify shortest paths in scientific social networks in order to make suggestions on the choice of how to reach another researcher or group of researchers at the lowest possible cost, even showing how the members of the network are connected process.*

**Resumo.** *As redes sociais científicas tem atraído atenção de pesquisadores que procuram entender sua evolução e seu comportamento. A quantidade de dados destas redes as torna fonte de informações valiosas. Este trabalho busca identificar menores caminhos em redes sociais científicas com o intuito de fazer sugestões no processo de escolha de como se chegar a um determinado pesquisador ou grupo de pesquisadores com o menor custo possível, apresentando ainda como os integrantes da rede estão conectados.*

**Keywords:** *Social Network Analysis, Complex Networks, Shortest Paths.*

## 1. Introdução

Para Sonnenwald (2007), a colaboração científica, referida no presente trabalho pode ser definida como uma interação entre um ou mais cientistas entre os quais o significado da realização de determinada tarefa é o mesmo e tal interação facilita a realização da mesma. Nessa rede, todos buscam obter e produzir maiores conhecimentos científicos.

Currículos da Plataforma Lattes são uma vasta fonte de informação para a criação e análise de redes sociais de pesquisadores devido a sua diversidade de área e quantidade de informações atualizadas pelos próprios pesquisadores.

O objetivo do trabalho consiste em encontrar a menor distância de um cientista até outro na rede de colaboração gerada a partir da análise dos currículos Lattes de um grupo de pesquisadores. Com o intuito de aperfeiçoar as relações entre autores e co-autores de artigos científicos, o presente trabalho visa apresentar a menor distância entre dois pesquisadores na rede.

## 2. Metodologia

Para o desenvolvimento do trabalho foi utilizado o maior repositório de currículos nacionais utilizado atualmente, a Plataforma Lattes. Para formação da rede científica foi considerada toda a base de dados da Plataforma com o auxílio de frameworks específicos para a extração de dados curriculares da Plataforma Lattes. (DIAS E MOITA, 2013)

Com auxílio de técnicas de descoberta de conhecimento em bases de dados (Knowledge Discovery in Databases – KDD), mais de 3,2 milhões de currículos foram obtidos e utilizados para o desenvolvimento do trabalho.

A partir dos dados extraídos, foi possível criar uma matriz de adjacência que representa com quais autores um determinado autor publicou algum artigo científico.

De posse desta matriz é possível identificar quais os autores estão vinculados ou isolados na rede em análise. E desta forma verificar quais as rotas para se chegar a um determinado pesquisador partindo de qualquer outro autor da rede, e diante disto se verificar qual o menor caminho entre pares de pesquisadores. Figura 1.

|                | 20000348934783 | 20003734983409 | 30002093238923 | 30000039457384 | 21000094937283 | 10902322200923 |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 20000348934783 | 0              | 1              | 0              | 1              | 0              | 0              |
| 20003734983409 | 1              | 0              | 1              | 0              | 0              | 1              |
| 30002093238923 | 0              | 1              | 0              | 1              | 0              | 0              |
| 30000039457384 | 1              | 0              | 1              | 0              | 1              | 0              |
| 21000094937283 | 0              | 0              | 0              | 1              | 0              | 1              |
| 10902322200923 | 0              | 1              | 0              | 0              | 1              | 0              |

**Figura 1. Tabela com identificação de relacionamentos**

Esta é uma tarefa relevante pois desta forma é possível identificar o quão distante dois elementos da rede se encontra e ainda trazer informações úteis sobre como dois elementos podem se relacionar e qual a melhor forma de se manter contato entre estes elementos por exemplo, para um trabalho em conjunto.

### 3. Referencial Teórico

Segundo Locke (2004) e Boaventura Netto (2003), o início dos estudos sobre a teoria dos grafos se deu por um matemático chamado Leonhard Euler (1707-1783), que buscava a possibilidade de percorrer, ou não, toda a cidade de Knigsberg (atualmente, Kaliningrado) cruzando uma de suas pontes só uma vez.

A teoria dos grafos busca uma explicação para todos os problemas da humanidade se baseando em objetos e nas ligações existentes, ou não, entre estes. Através desse método é possível usar algumas formas de representação para melhor estudo dos grafos. Segundo Scott (2004), muitas características fundamentais das redes sociais podem ser analisadas por meio da manipulação direta de matrizes. Uma destas é chama de matriz de adjacência e consiste em criar uma matriz de  $V \times V$  adicionando peso 1 em todos os campos que representarem ligações entre dois vértices. Figura 2.

|   | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 1 | 0 | 1 |
| B | 1 | 0 | 1 | 0 |
| C | 0 | 1 | 0 | 1 |
| D | 1 | 1 | 1 | 0 |

**Figura 2. Matriz de Adjacência**

A Figura 2 representa uma matriz adjacência, onde são representadas as ligações entre os vértices A, B, C e D. É possível verificar que a matriz diagonal está preenchida com 0, pois, nesse caso, não é possível existir uma aresta laço. Os outros campos estão preenchidos com peso 1 onde existe a ligação entre os vértices.

Em alguns casos os vértices e as arestas correspondem a objetos físicos do mundo real, em outros, os vértices são objetos reais enquanto as arestas correspondem a relacionamentos intangíveis, e ainda existem casos em que vértices e arestas são puras abstrações [Kleinberg e Tardos, 2006].

Com a utilização de técnicas originadas da teoria dos grafos também é possível buscar o menor caminho de um vértice até qualquer outro vértice. Isso é útil para se determinar com quantos passos um determinado pesquisador consegue chegar aos outros pesquisadores da rede.

### **3.1 Busca em Menor Caminho**

Em sua essência, a busca de caminho é simplesmente o processo de movimento de um local inicial ao destino desejado [Bourg & Seeman, 2004].

Existem muitas técnicas para a descoberta de menores caminhos, uma pode se sobressair em certas circunstâncias, mas se comportar de maneira severamente errada em outras. O sucesso de uma dada abordagem depende dos requerimentos e às suposições do problema. Quanto mais fatores tenham de ser levados em consideração, maior será a complexidade da solução. A busca de caminho, entretanto, é complexa.

Existem diversas formas de encontrar o menor caminho entre os vértices de determinado vértice. Quando se busca estabelecer essa menor distância é necessário avaliar uma série de quesitos e estudar qual das técnicas será mais bem implementada.

Uma das técnicas amplamente utilizadas é a do algoritmo de Dijkstra [Dijkstra, 1959]. Este método, ao contrário da Busca em Largura (BFS), apresenta sempre o melhor caminho entre dois vértices, desde que não existam pesos negativos em suas arestas. Outro ponto positivo é que na publicação original deste algoritmo, ocorre a busca desinformada pelo menor caminho, ou seja, é encontrada a menor distância de um vértice até todos os outros vértices do grafo.

Para que o algoritmo seja executado é necessário que sejam repassados dois parâmetros: o pesquisador de origem e o grafo a ser pesquisado. Partindo da matriz de adjacência, é possível percorrer o grafo e dizer qual a menor distância deste pesquisador até todos os outros pesquisadores do grafo.

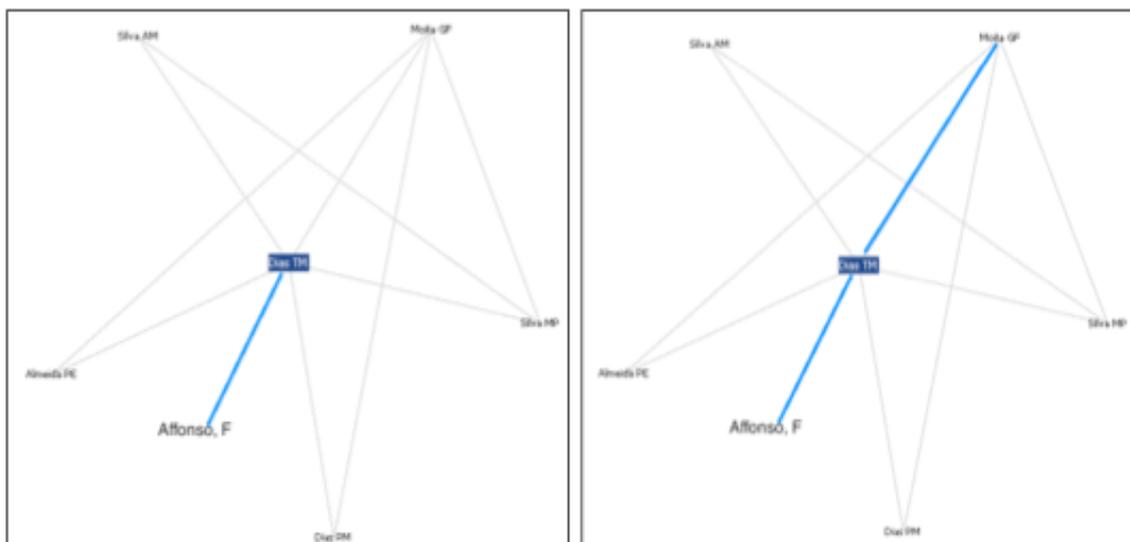
## **4. Desenvolvimento**

A tarefa de formar um grupo de pesquisa produtivo é um processo complexo. Verificar cada currículo e analisar se um usuário faz pesquisa com o mesmo tema de outros pesquisadores não é uma tarefa trivial. Diante disto, com a rede formada pelos pesquisadores é possível identificar pesquisadores que possuem vínculos mais próximos de outros pesquisadores e desta forma facilitar a identificação de possíveis colaboradores.

Outro fator importante na análise de redes sociais científicas é a realização da busca de menores caminhos entre instituições de ensino. Com isto, é possível que pesquisadores verifiquem como a sua instituição está vinculada com outras instituições e qual o menor caminho para se estabelecer algum tipo de contato.

Para a implementação das tarefas descritas, foi desenvolvido uma interface gráfica para apresentar a busca por menores caminhos em tempo real para os usuários. Com isto é possível, por exemplo, escolher um possível orientador e encontrar qual o menor caminho para se estabelecer um contato mais consistente.

Com o auxílio da matriz de adjacência gerada e dos algoritmos de busca implementados, é possível mostrar quantos passos são necessários para chegar a um determinado pesquisador ou instituição de ensino de interesse. Um exemplo desse processo é apresentado na Figura 3.



**Figura 3. Representação de menor caminho em uma rede de colaboração científica**

É possível observar na rede gerada, como um pesquisador está vinculado a outro de forma interativa, considerando o menor custo possível, ou seja, a menor quantidade de saltos necessários.

O trabalho consiste em realizar estas tarefas de forma automatizada e em tempo real, apresentando qual a distância de um pesquisador até todos os outros pesquisadores da rede. Este processo irá facilitar a tarefa de se estabelecer contatos e facilita a localização entre pesquisadores integrantes da rede.

## 5. Conclusões

O trabalho em colaboração é fator cada vez mais importante para a evolução das pesquisas. O momento atual é de uma sociedade em rede amplamente conectada e necessitamos cada vez mais, estar conectados uns com os outros. Por isso, é importante identificar quais elementos são mais influentes ou mais importantes em uma rede social, principalmente em uma rede científica. Por isso, o trabalho se mostra relevante, já que com os resultados obtidos é possível identificar colaboradores com maior nível de importância na rede, baseado em seu grau.

A busca por orientandos ou orientadores é extremamente difícil. Existe um grande impasse no momento de escolher a pessoa certa para ser um orientador tendo em vista a dificuldade de se estabelecer um contato mais eficiente. Com a análise da rede científica formada pela Plataforma Lattes essa decisão é facilitada apresentando qual o menor caminho entre o interessado e o possível orientador.

Outro fator muito importante em desenvolvimento é a realização da busca em menor caminho entre Instituições de ensino. Com isto, os pesquisadores poderão verificar se a sua instituição já realizou algum tipo de pesquisa com outras instituições e qual o vínculo entre elas.

## 6. Referências

BOAVENTURA NETTO, P. O. Grafos: teoria, modelos, algoritmos. 3. ed. São Paulo: Edgard Blucher, 2003.

BOURG, David M.; SEEMANN, Glenn. AI for Game Developers. North Mankato: O'Reilly Media, Inc., 2004.

T. M. R. Dias and G. F. Moita, Extração e Modelagem de Redes de Colaboração Científica. In: Conferência IADIS Ibero-Americana WWW/Internet, 2013, Porto Alegre. Brazil.

DIJKSTRA, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1 (1), 269-271.

KLEINBERG, J, Tardos, É. *Algorithm desing*. Pearson Education, 2006. ISBN: 0-321-29535-8.

LOCKE, S. C. *Graph theory*. 2004. Disponível em: <[www.math.fau.edu/locke/graphthe.htm](http://www.math.fau.edu/locke/graphthe.htm)>. Acesso em: 31 jan. 2005.

SONNENWALD, D. H. Scientific collaboration. *Annual Review of Information Science and Technology*, v.41, n.1, p.643-81. 2007.

SCOTT, J. *Social network analysis: a handbook*. 2004. Não-publicado. Disponível em: <[www.analytictech.com/mb119/tableof.htm](http://www.analytictech.com/mb119/tableof.htm)>. Acesso em: 8 dez. 2004.