

## Association for Information Systems AIS Electronic Library (AISeL)

---

Proceedings of the XI Brazilian Symposium on  
Information Systems (SBSI 2015)

Brazilian Symposium on Information Systems  
(SBIS)

---

5-2015

# Analysis of Scientific Collaboration Network as a Management Tool for Graduate Programs

Aurélio Ribeiro Costa

*Universidade de Brasília*, [arcosta@gmail.com](mailto:arcosta@gmail.com)

Célia Ghedini Ralha

*Universidade de Brasília*, [ghedini@cic.unb.br](mailto:ghedini@cic.unb.br)

Follow this and additional works at: <http://aisel.aisnet.org/sbis2015>

---

### Recommended Citation

Costa, Aurélio Ribeiro and Ralha, Célia Ghedini, "Analysis of Scientific Collaboration Network as a Management Tool for Graduate Programs" (2015). *Proceedings of the XI Brazilian Symposium on Information Systems (SBSI 2015)*. 64.  
<http://aisel.aisnet.org/sbis2015/64>

This material is brought to you by the Brazilian Symposium on Information Systems (SBIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the XI Brazilian Symposium on Information Systems (SBSI 2015) by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Análise de Rede de Colaboração Científica como Ferramenta na Gestão de Programas de Pós-graduação

Alternative Title: Analysis of Scientific Collaboration Network as a Management Tool for Graduate Programs

Aurélio Ribeiro Costa  
Departamento de Ciência da Computação  
Instituto de Ciências Exatas  
Universidade de Brasília  
Brasília - DF - Brasil  
arcosta@gmail.com

Célia Ghedini Ralha  
Departamento de Ciência da Computação  
Instituto de Ciências Exatas  
Universidade de Brasília  
Caixa Postal 4466 - Brasília - DF  
CEP 70.904-970  
ghedini@cic.unb.br

## RESUMO

O desempenho de um programa de pós-graduação é aferido pela CAPES, em parte pelo seu nível de publicação, para tanto, faz-se necessário que a gestão do programa seja feita de forma a maximizar a qualidade das publicações que são realizadas pelos pesquisadores associados. No contexto dos relacionamentos de coautoria em publicações, a análise de rede de colaboração científica se mostra uma ferramenta adequada para avaliar as parcerias já formadas, bem como para estimular a formação de novas parcerias. Neste artigo é apresentada uma ferramenta de análise de rede aplicada à gestão de um programa de pós-graduação de uma Universidade Federal. Os dados coletados do programa foram modelados em um banco NoSQL orientado a grafo, incluindo todas as publicações dos docentes vinculados. Ressaltamos a utilidade ao gestor do programa, bem como aos demais pesquisadores, do módulo de recomendação de parcerias. O desenvolvimento da pesquisa utilizou a metodologia *Design Science Research* para guiar tanto a construção do artefato quanto para elaborar a documentação associada. Preliminarmente, utilizando dados de um programa de pós-graduação, pode-se perceber o potencial de recomendação de novas parcerias de co-autoria na rede de colaboração previamente formada.

## Palavras-Chave

Sistemas de Recomendação, NoSQL

## ABSTRACT

The performance of a graduate program is assessed by CAPES partly by their level of publication. Therefore, it is necessary that the program chair has instruments to analyze the quality of publication produced by the associated re-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2015, May 26th-29th, 2015, Goiânia, Goiás, Brazil  
Copyright SBC 2015.

searchers. In the context of co-authoring relationships in publications, network analysis has been proved to be an appropriate tool to evaluate the relationships already formed, and to stimulate the formation of new relationships. This paper presents a network analysis tool applied to real data of a graduate program from a Brazilian Federal University. The data was modeled in a NoSQL graph oriented database including all associated researchers' publication. We emphasize the usefulness of the partners' recommendation module to the program chair, as well as associated researchers. The development of this research work used the *Design Science Research* methodology to guide both the construction of the artefact and the associated documentation. Preliminary, using data from a graduate program, we can note the recommendation potential to integrate new partners to the scientific network already formed.

## Categories and Subject Descriptors

H.2.8 [Database Applications]: Scientific databases; H.4.2 [Types of Systems]: Decision support; J.1 [ADMINISTRATIVE DATA PROCESSING]: Education

## General Terms

Social Network Analysis, Design Science Research

## Keywords

Recommending Systems, NoSQL

## 1. INTRODUÇÃO

Um dos fundamentos da gestão de um programa de pós-graduação baseia-se na maximização dos indicadores de produção acadêmica, especialmente a publicações em periódicos bem qualificados com abrangência internacional. Os indicadores de produção científica e acadêmica são elementos fundamentais nas avaliações realizadas pela CAPES<sup>1</sup>, fundação vinculada ao Ministério da Educação responsável por avaliar os programas de pós-graduação em todos os estados da Federação. Nesse sentido, faz-se necessário que os gestores desses programas disponham de uma ferramenta que os

<sup>1</sup><http://www.capes.org.br>

permita visualizar a distribuição das publicações que foram realizadas pelos pesquisadores. O objetivo de tal ferramenta é auxiliar no direcionamento de ações objetivando a melhoria dos indicadores de produção científica e acadêmica.

Este trabalho objetiva apresentar uma ferramenta de auxílio à gestão dos indicadores de produção acadêmica de um programa de pós-graduação. A abordagem utilizada para construção dessa ferramenta foi baseada na análise da rede de colaboração científica formada pelos pesquisadores vinculados ao programa. Também é objetivo deste trabalho detalhar a implementação da funcionalidade de recomendação de parcerias nas futuras publicações a serem desenvolvidas pelos pesquisadores vinculados ao programa.

As demais seções desse trabalho estão distribuídas da seguinte forma: na Seção 2 são apresentados os trabalhos que de alguma forma contribuíram para o desenvolvimento desse trabalho de pesquisa; na Seção 3 são apresentados os fundamentos teóricos que deram base para a pesquisa; na Seção 4 a proposta de solução é descrita; na Seção 5 é realizada uma análise dos dados das publicações presentes nos currículos Lattes<sup>2</sup> dos pesquisadores vinculados ao programa, limitando-se às publicações realizadas nos últimos dez anos; na Seção 6, a ferramenta de recomendação de parcerias em publicações é detalhada; e finalmente, na Seção 7 são apresentadas as conclusões e os trabalhos futuros.

## 2. TRABALHOS RELACIONADOS

Como trabalhos correlatos destacamos o realizado em [1], o qual apresenta uma ferramenta para análise das informações de publicação dos pesquisadores vinculados ao Instituto Nacional de Ciência e Tecnologia para Controle do Câncer (INCTCC). Neste trabalho foi desenvolvido um ambiente no qual seja possível a análise de produção científica através de métricas orientadas a grafos como grau médio, *betweenness* e *closeness*, no entanto, não implementa a funcionalidade de recomendação de parcerias, o que a diferencia do presente trabalho. No desenvolvimento de tal ferramenta foi utilizada a abordagem de *Business Intelligence*, com os dados persistidos em um banco de dados relacional Microsoft SQL Server e as consultas no grafo de relacionamentos extraídas pela ferramenta Gephi[2]. Na implementação realizada no presente trabalho, as consultas ao modelo de dados são geradas diretamente no banco de dados orientado a grafo, o que evita a tradução das consultas entre o modelo orientado a grafo e o modelo relacional diminuindo assim a operabilidade das consultas.

O trabalho realizado em [3] propõe uma modificação na plataforma de rede de colaboração utilizada pelo C.E.S.A.R. de forma a dotá-la de um componente de recomendação de conteúdo/pessoas. Foi desenvolvida uma ferramenta chamada SWEETS 1.0 que serviu como motivação inicial para a implementação do módulo de recomendação apresentado no presente trabalho. Embora o SWEETS 1.0 seja a implementação de um sistema de recomendação, a estratégia de filtragem baseada em conteúdo usada por ele difere da estratégia de filtragem colaborativa baseada em vizinhança empregada no presente trabalho.

O trabalho realizado em [4] aplica a metodologia *Design Science Research*, assim como neste trabalho, porém com uma finalidade distinta pois concentrou-se na construção de

um artefato para calcular o tamanho ideal de participantes de um bate-papo no contexto de educação a distância. Sua contribuição para este trabalho se dá pela aplicação do *Design Science Research* em uma pesquisa em Sistemas de Informação, uma vez que o uso do *Design Science Research* guiou a produção de um modelo matemático em [4], já no presente trabalho o objetivo é a construção de um artefato de *software*.

Destacamos também o trabalho realizado por [5], o qual introduz várias métricas para a análise de redes de colaboração científica buscando identificar grupos de pesquisadores com interesses em comum no desenvolvimento de suas pesquisas. Para tanto, foi construída uma rede de colaboração científica multi-camada, onde cada camada representa um tipo de relacionamento – coautoria, participação em projeto, orientação e produção técnica são os relacionamentos abordados. É através dessa rede que são realizadas análises como a aplicação de algoritmos de *clustering* para buscar os grupos de pesquisadores. No trabalho é apresentada ainda a métrica de quantidade de relacionamentos em comum, ponderadas temporalmente, que serviu de motivação para o cálculo do fator de recomendação apresentado neste trabalho.

## 3. MÉTODOS E TECNOLOGIAS

Alguns métodos e tecnologias foram utilizados nesta pesquisa. Nesta seção será descrito brevemente cada um deles para que seja possível relacioná-los na proposta de solução apresentada na Seção 4.

### 3.1 Redes sociais e de colaboração científica

Segundo [5], uma rede é um conjunto de objetos conectados entre si. Pode-se representar uma rede por um grafo, onde os nós ou vértices estão relacionados/conectados por arestas. Uma rede social reflete uma estrutura social, a qual pode ser representada por indivíduos ou organizações e suas relações. Em geral, as relações representam um ou mais tipos de interdependência – como ideias ou religiões, ou relacionamentos mais específicos – como troca de conhecimento/informação ou amizade. Através dessa estrutura social é possível estudar a troca de informação entre indivíduos e organizações[5].

Redes de colaboração científica são redes sociais cujos relacionamentos representam algum tipo de interação científica. Segundo [6], uma rede social pode ser representada por um grafo  $G = (V, E)$ , onde  $V$  denota um conjunto de indivíduos em consideração e  $E$  um conjunto de relacionamentos existentes entre esses indivíduos. Uma rede social pode suportar um ou mais tipos de relacionamentos, como amizade, paternidade ou colaboração profissional. Intuitivamente, relacionamentos possuem diferentes intensidades que refletem a força das ligações sociais. A intensidade de um relacionamento é usualmente representada por uma função  $w(e)$  com  $e \in E$ , que associa um peso para aresta na rede.

### 3.2 Design Science Research

O paradigma *Design Science Research* é adotado nessa pesquisa por possibilitar a construção de um artefato que materialize as conjecturas teóricas e cuja avaliação permita refutar ou comprovar as teorias iniciais estabelecidas na pesquisa. De acordo com [7], *Design Science Research* constitui uma outra “lente”, um conjunto de técnicas com perspectivas analíticas e sintéticas, para realização de pesquisa em

<sup>2</sup>Plataforma Lattes do CNPQ. Disponível em: <http://lattes.cnpq.br>

sistemas de informação.

Nesse contexto, *Design Science Research* envolve a geração de conhecimento novo através do desenvolvimento de artefatos inovadores e da análise do uso e/ou do desempenho de tais artefatos por meio de reflexão e abstração. Tais artefatos incluem, mas não se limitam a, interfaces homem/máquina, algoritmos, metodologias de projeto de sistemas e linguagens. Esta pesquisa tem o objetivo de construção de um artefato que modele as parcerias entre pesquisadores através de relações de coautoria em publicações científicas.

Uma conjectura teórica adotada neste trabalho é fundamentada na hipótese de que pesquisadores com uma boa rede de colaboração, seja em tamanho ou em qualidade, estejam aptos a ter uma boa produção científica qualificada. Buscando validar ou refutar essa conjectura, foi construído um grafo de relacionamento baseado na coautoria entre pesquisadores.

Para a construção do grafo de relacionamento, construído inicial do *Design Science*, foram adotados dois tipos de nós, o nó pesquisador e o nó publicação. Também foi definido um tipo de aresta, o qual representa a relação de autoria de uma publicação. O ponto de partida para compreender a relação de parceria entre pesquisadores foi realizado através da coleta de dados dos currículos Lattes dos professores vinculados ao Departamento de Ciência da Computação da Universidade de Brasília e atuantes no programa de pós-graduação em Informática, os quais se relacionam através de autoria em artigos completos publicados em periódicos.

Embora as publicações em conferências nacionais e internacionais também estejam presentes nos currículos Lattes dos pesquisadores, optou-se por não usar esse item pela baixa confiabilidade dos dados inserido. Todavia, é possível ainda a inclusão desses dados com a configuração da ferramenta de extração.

As relações de colaboração foram extraídas através dos autores dos artigos cadastrados na plataforma Lattes. Foi utilizado como chave de acesso no banco de dados os nomes dos autores dos artigos e os nomes utilizados para referência de cada autor na Seção Identificação do currículo Lattes. Assim, com os nomes dos autores dos artigos e os nomes utilizados para referência de cada autor, foi possível a geração das arestas, representando as autorias do grafo permitindo então a visualização das parcerias já realizadas.

### 3.3 Bancos de dados NoSQL

Com o desenvolvimento de tecnologias de armazenamento de dados NoSQL, as quais buscam formas alternativas à abordagem relacional para problemas específicos, surgiu o interesse em não apenas modelar uma rede social como um grafo, mas também em armazená-la dessa forma[8].

Neste trabalho, foram estudadas diversas soluções de banco de dados NoSQL e optou-se por empregar o *Banco de dados em Grafo Neo4j*[9]. O Neo4j possui licença GPL, com extensa documentação disponível *on-line*, ao contrário de soluções como OrienteDB[10], AragoDB[11], VelocityGraph[11]. Essas implementações de NoSQL orientado a grafo utilizam armazenamento baseado em documento ou *Resource Description Framework*, RDF[12] (metamodelo para armazenamento de informações implementado em recursos web pelo consórcio W3C). Foram estudados ainda os bancos de dados relacionais Oracle, com sua biblioteca Spatial[13], e PostgreSQL, com a extensão PostGIS [14], porém usando o para-

digma relacional para armazenamento dos dados em tabelas essas alternativas apenas simulam o modelo em grafo por meio de *stored procedures*. Essa abordagem aumenta a operacionalidade, uma vez que é necessário realizar uma tradução da consulta no grafo para uma consulta relacional.

### 3.4 Sistemas de Recomendação

Sistemas de recomendação diferem quanto a forma que eles analisam os dados para desenvolver noção de afinidade entre usuários e itens que pode ser usada para identificar o casamento de pares usuário-item. Existem duas formas mais difundidas de sistemas de recomendação, a filtragem colaborativa, a qual analisa apenas interações históricas e a filtragem baseada em conteúdo, que usa os dados disponíveis nos perfis de usuários e nos atributos dos itens. Há ainda técnicas híbridas que combinam características de ambas as abordagens [15].

Conforme [5], recomendação de parcerias, ou predição de link no contexto de um grafo social, pode ser realizada de duas diferentes formas, porém complementares. A primeira examina um sociograma desbalanceado de estruturas sociais buscando transformá-lo em uma estrutura social balanceada. Como exemplo, se duas pessoas possuem muitos amigos em comum então essas duas pessoas devem se conhecer, em uma abordagem usando filtragem colaborativa. A segunda abordagem analisa o conteúdo da comunicação entre os indivíduos, buscando por indivíduos que compartilham o mesmo interesse. Como exemplo, se dois pesquisadores publicam sobre um mesmo assunto então eles deveriam se encontrar, nesse caso o método mais adequado para o sistema de recomendação é a filtragem baseada em conteúdo.

Segundo [15], o objetivo de um sistema de recomendação é gerar recomendações significativas a um conjunto de usuários para itens ou produtos que lhe possa interessar. Sugestões de livros na Amazon, amigos no Facebook ou filmes no Netflix são exemplos reais do uso de sistemas de recomendação na indústria. O projeto de um sistema de recomendação depende do domínio de características particulares dos dados disponíveis. Tais dados podem armazenar a relevância de um determinado item para um usuário, tal quando se classifica um item com estrelas em uma loja virtual ou quando se aponta o gostar de um comentário em uma rede social. Adicionalmente, o sistema pode ter acesso a atributos específicos de usuários ou itens como informações demográficas e descrição do item respectivamente.

Nesse trabalho, foi adotada uma abordagem híbrida combinando a filtragem colaborativa baseada em vizinhança através da similaridade entre os pesquisadores e a filtragem baseada em conteúdo usando os dados históricos das publicações.

## 4. PROPOSTA DE SOLUÇÃO

A proposta de um artefato para análise da rede de colaboração científica foi iniciada com a modelagem das informações contidas nos currículos Lattes dos pesquisadores, mais especificamente dos dados referentes às publicações em periódicos. Em seguida, foi definido o fluxo de tratamento desses dados desde a extração na plataforma Lattes até a visualização da informação.

### 4.1 Modelo de dados

O modelo de dados utilizado na solução é apresentado na Figura 1, a qual destaca as entidades e seus atributos armazenados no grafo da rede de colaboração científica modelada.

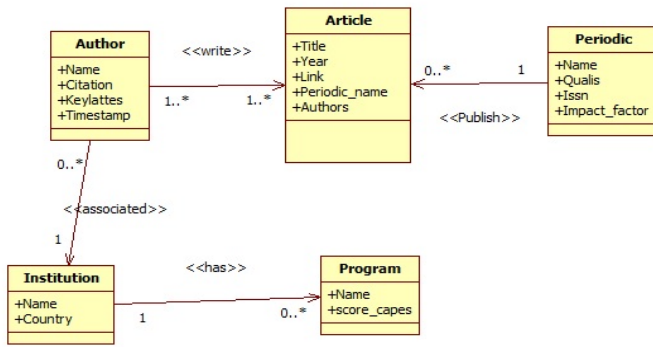


Figura 1: Diagrama de classes representando o modelo de dados

## 4.2 Processo metodológico da solução

A Figura 2 apresenta o processo metodológico da proposta, incluindo desde a coleta dos nomes dos professores até a visualização do grafo. Note que o primeiro elemento é o sub-processo de pré-processamento, o qual inclui três atividades:

- busca lattes – realiza a busca do currículo Lattes do autor;
- trata identificação de autoria - extrai do currículo Lattes os nomes usados pelo autor em citações;
- captura artigos completos publicados em periódicos – extrai as publicações que serão usadas posteriormente para construção da rede de colaboração.

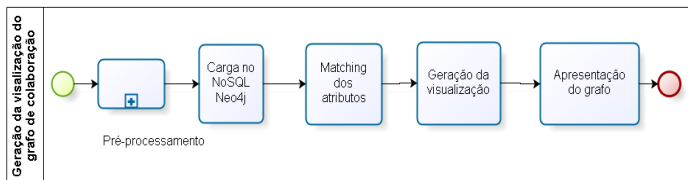


Figura 2: Processo de carga e visualização

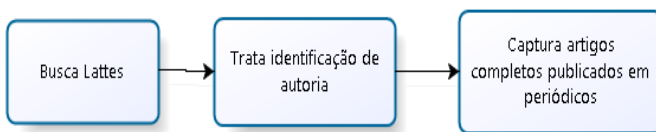


Figura 3: Sub processo de pré processamento

Após o sub-processo de pré-processamento, detalhado na Figura 3, a tarefa de carga no Neo4j é iniciada, quando será realizado o *matching* dos atributos de autoria para criação da rede de colaboração. Na sequência, ocorrem a geração da visualização, quando são construídos os gráficos. Posteriormente é apresentado o grafo para o usuário.

## 4.3 Arquitetura da solução

A implementação do artefato iniciou com a construção de um *script* para realizar uma busca dos dados do currículo Lattes de cada pesquisador vinculado ao programa de pós-graduação, salvando os dados da carga em arquivos de texto para possibilitar um pré-processamento antes da carga no banco Neo4j. Durante essa fase, algumas medidas corretivas precisaram ser adotadas, como o tratamento de homônimos, abreviações e inconsistências léxicas nos dados de origem. Com o resultado da busca de currículos, foram obtidos os dados de identificação dos pesquisadores, assim como, as publicações deixando essas entidades, autores e publicações, ainda sem definição de relacionamento.

Procedeu-se a carga do grafo com os dados de autores e de publicações obtidos na etapa anterior para então realizar o casamento entre as informações de identificação de cada autor como as informações de autoria de cada publicação. Na sequência, foi construído um artefato de visualização, o qual foi implementado utilizando o padrão de projeto MVC – Model-View-Controller[16]. Os elementos foram implementados da seguinte forma:

- *View* – desenvolvido na linguagem Javascript utilizando a biblioteca D3[17], adequada para visualização de dados;
- *Controller* – responsável por manipular os objetos do grafo para serem apresentados ao elemento de visualização, utilizando o *framework* web2py[18], escrito na linguagem Python;
- *Model* – implementado através do banco de dados em grafo Neo4j, o qual define a forma de persistência dos objetos de dados.

Uma das vantagens em utilizar um banco de dados orientado a grafo nesse contexto é a facilidade em realizar consultas no banco, como se estas estivessem sendo realizadas no próprio modelo conceitual. Os experimentos realizados utilizaram uma massa de teste com 401 vértices (entre autores e publicações) e 376 arestas, o que não apresentou problema de desempenho no grafo gerado.

Uma limitação identificada no decorrer do desenvolvimento do artefato foi a atualização dos dados de publicação nos currículos dos pesquisadores. Bem como, o campo de nomes usados em citações para que haja a correta análise das informações, uma vez que citações constituem a base para o relacionamento entre autores e publicações. A presença de homônimos na base dos currículos Lattes também foi uma dificuldade, a qual foi superada inicialmente através de intervenções manuais através da inclusão dos currículos que a ferramenta não conseguiu carregar. Claro que estas intervenções foram viáveis tendo em vista a pequena quantidade de currículos processados, 48 no total. Houve também dificuldade no processamento de caracteres latinos, como letras acentuadas, o que gera a necessidade de uma maior atenção na codificação de caracteres utilizados, para tanto foi necessário realizar uma normalização dos caracteres acentuados reduzindo-os aos caracteres presentes no conjunto de caracteres ASCII<sup>3</sup>.

<sup>3</sup>Sigla em inglês para *American Standard Code for Information Interchange* (Código Padrão Americano para o Intercâmbio de Informação)

Um outro aspecto que deve ser considerado é que embora já existam ferramentas para auxiliar a realização de análise de redes sociais, como Pajek[19], NVIVO[20] e UCINET[21], optou-se pela não utilização dessas ferramentas uma vez que nenhuma delas possui a capacidade de consultar diretamente o banco de dados em grafo, no qual os dados são persistidos. O uso destas ferramentas traz a necessidade de desenvolver um artefato para organizar os dados de forma que as ferramentas pudessem tratar. Dessa forma, o esforço necessário foi direcionado para a exploração da capacidade de realizar a análise de um modelo de rede de colaboração já persistido em um grafo. Salienta-se que os construtos implementados representam o ponto de partida para o desenvolvimento da pesquisa cujo objetivo é a construção de um modelo de recomendação que utilizará os dados persistidos no banco de dados em grafo já implementado.

### 5. ANÁLISE DOS DADOS DE PUBLICAÇÃO

Utilizando os dados dos pesquisadores vinculados ao programa de Pós-Graduação em Informática da universidade XXXX, bem como as métricas definidas em [5], foi obtido o conjunto de valores que caracterizam a rede de colaboração científica. Desta forma, é possível a utilização de técnicas de filtragem de informação que viabilizem a construção de um modelo para realização de recomendações de novas parcerias. Nos últimos dez anos as publicações seguiram o padrão demonstrado na Figura 4, que exibe graficamente as informações da Tabela 1.

Na Tabela 1, é detalhada a evolução na quantidade de publicações como também a quantidade de coautorias internas, ou seja entre autores do mesmo programa de pós-graduação, e coautorias externas, na qual a relação de coautoria é realizada com um autor não vinculado ao programa de pós-graduação. É possível verificar na coluna *Fator de coautoria* a relação entre os relacionamentos de coautoria internos e externo, o qual foi calculado usando a Equação 1, para cada ano:

$$FC_i = \frac{\sum_{a \in A} (numAutInter_a - numAutExt_a)}{\sum numPublicacoes_i} \quad (1)$$

onde *numAutInter* representa a quantidade de autorias internas, *numAutExt* a quantidade de autorias externas, *numPublicacoes* a quantidade total de publicações em um determinado ano e *A* é o conjunto de autores de uma determinada publicação. Note que quanto mais negativo for o valor de FC, mais coautores externos existem nas publicações, demonstrando maior interação com grupos de pesquisa externos ao programa.

Na Figura 4, o tracejado representa o total de publicações realizadas, os triângulos as autorias entre os pesquisadores do departamento, os círculos as autorias externas e a linha contínua o fator de coautoria que para melhorar a visualização foi multiplicado por dez. Na mesma figura é possível identificar que nos anos de 2005 e 2013 houve um aumento das publicações com um expressivo incremento na participação de autores externos. Se em 2013 houver uma maior quantidade absoluta de publicações, por outro lado, no ano de 2005 houve uma maior participação relativa de autores externos, o que gerou uma minimização no *Fator de coautoria*. Por outro lado, no ano de 2007, percebe-se uma baixa interação com coautores externos, o que reflete a baixa quantidade de publicações realizadas, 15, número esse que

Tabela 1: Quantidade de publicações

Ano	Publicações	Coautorias internas	Coautorias externas	Fator de coautoria
2004	15	17	38	-1.4
2005	26	27	81	-2.0769
2006	19	22	45	-1.2105
2007	15	22	24	-0.1333
2008	26	29	62	-1.2692
2009	20	23	48	-1.25
2010	20	26	43	-0.85
2011	19	23	48	-1.3157
2012	21	27	51	-1.1428
2013	36	39	99	-1.6666
2014	25	29	56	-1.08

também ocorre em 2004 porém neste ano a quantidade de coautorias externas é aproximadamente 50 % maior que em 2007, o que eleva o FC em 10 vezes em 2004 na comparação com 2007.

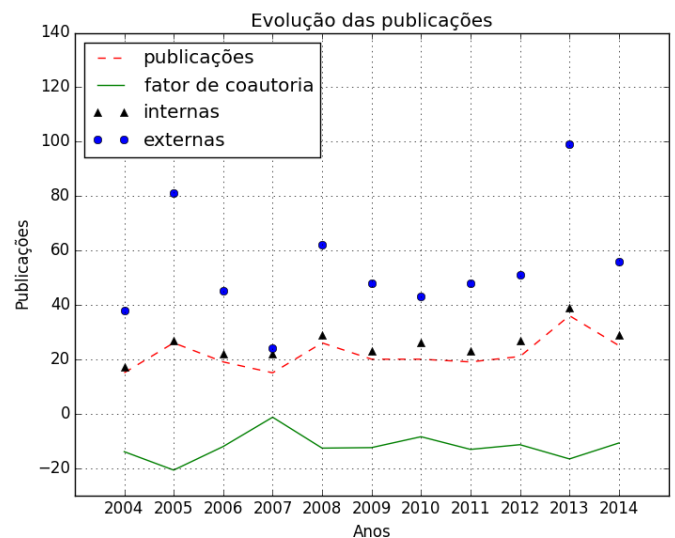


Figura 4: Publicações por ano

### 6. RECOMENDAÇÃO DE PARCERIAS NA AUTORIA DE PUBLICAÇÕES

Um aspecto identificado na análise dos dados é que a formação de parceria na coautoria de publicações melhora os indicadores de produção, o que corrobora com a conjectura teórica definida na Seção 3.2, a qual defende que pesquisadores com uma boa rede de colaboração, seja em maturidade ou em qualidade, estejam aptos a ter uma boa produção científica qualificada. Nesse sentido, foi desenvolvido um módulo de recomendação para auxiliar a identificar possíveis parcerias na realização de pesquisas.

Em virtude da modelagem dos dados utilizada neste trabalho ser centrada no relacionamento entre pesquisadores e publicações em periódicos, conforme Seção 4.1, a metodologia de filtragem colaborativa baseada em vizinhança se

mostrou adequada, uma vez que nela um subgrupo de usuários são escolhidos baseado na similaridade desses usuários com um usuário ativo previamente selecionado. Já a combinação ponderada das relações usuário/item – pesquisador/publicação –, utilizando a abordagem baseada em conteúdo, compõem técnicas usadas para inferir predições para o usuário ativo, dessa forma caracterizando uma abordagem híbrida em sistemas de recomendação.

No contexto desse trabalho, os itens a serem considerados são os periódicos indexados, enquanto a ponderação dos pares pesquisador/periódico será baseada no decurso de tempo entre o ano da publicação e o ano atual em uma abordagem similar à utilizada por [5]. Assim, quanto mais publicações dois pesquisadores possuírem em um mesmo periódico maior será a similaridade entre eles, mesmo que eles nunca tenham realizado uma publicação em conjunto. O Fator de influência para cada publicação é dado pela Equação 2:

$$\rho_i = \frac{1}{e^{BY-PY}} \quad (2)$$

onde  $BY$  é o ano no qual a análise foi realizada, no caso 2014 e  $PY$  é o ano de publicação do trabalho no periódico. Dessa forma é possível calcular as recomendações conforme o somatório da Equação 3:

$$Rec = \sum_{i=1}^n \rho_i \quad (3)$$

onde  $n$  representa o número de publicações do autor ativo.

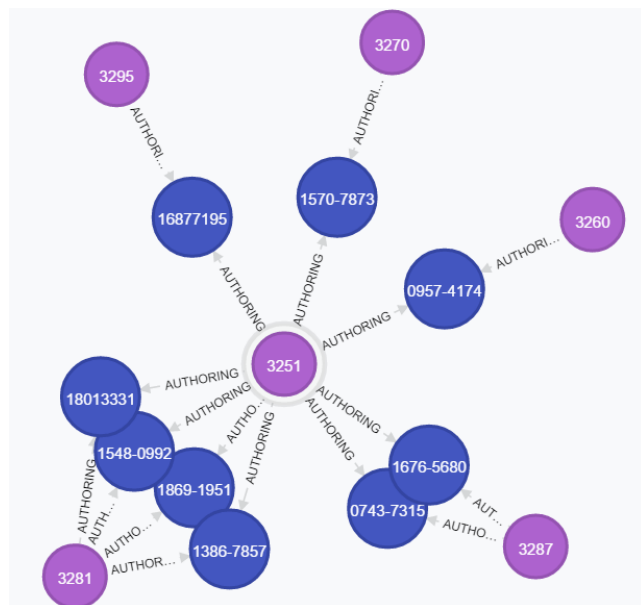
Escolhendo aleatoriamente como usuário ativo aquele com código de identificação 3251, foi gerado um ranking com os nove autores cujas probabilidades do usuário ativo formar parceria são maiores, conforme apresentado na Tabela 2, em ordem decrescente do índice de recomendação. Foram exibidas apenas nove recomendações uma vez que a partir da décima o índice é igual a zero.

**Tabela 2: Recomendações de parcerias para o pesquisador 3251**

#	Id	Rec
1	*3281	0.4193727265647525
2	*3270	0.36787944117144233
3	*3260	0.36787944117144233
4	3271	0.050122530995766454
5	*3295	0.049787068367863944
6	*3287	0.001292744523219513
7	3275	0.00033546262790251185
8	3284	0.00033546262790251185
9	3293	0.00033546262790251185

O ranking de recomendações apresentado na Tabela 2 contempla apenas pesquisadores vinculados ao programa de pós-graduação objeto dessa análise, uma vez que apenas estes tiveram seus currículos Lattes carregados no grafo de co-autorias. Ressalta-se no entanto que a mesma abordagem pode ser utilizada com pesquisadores de outros programas de diversas universidades, ou mesmo de fora do país. Ainda com relação a Tabela 2, os pesquisadores cujas identificações foram marcadas com asterisco já realizaram pelo menos uma publicação em conjunto com o usuário ativo, o que pode ser observado através da Figura 5 que apresenta a rede de colaboração já formada pelo usuário ativo. Nela, os círculos em lilás representam os pesquisadores identificados pelo índice

do nó no grafo, já os círculos em azul representam os periódicos usados pelos pesquisadores identificados pelo ISSN.



**Figura 5: Rede de colaboração centrada um usuário aleatório**

Analisando os dados da Tabela 2 em conjunto com a Figura 5 percebe-se que o usuário 3281, por possuir a maior quantidade de publicações realizadas em conjunto com o usuário ativo, aparece em primeiro no ranking. No entanto, como a quantidade de publicações é ponderada pela antiguidade de cada publicação no cálculo do índice de recomendação, o usuário 3287, que possui a segunda maior quantidade de publicações em conjunto com o usuário ativo aparece em sexto lugar no ranking. Assim, dentre as nove recomendações realizadas pelo módulo de recomendação, cinco delas são no sentido de reforçar parcerias já existentes e as outras quatro são criação de novas. Inclusive, com uma parceria nova (usuário 3271) possuindo índice de recomendação maior que duas outras parcerias já existentes.

Nas três últimas linhas da Tabela 2 verifica-se um empate entre os pesquisadores recomendados, isso se deve ao fato dos três últimos pesquisadores terem realizado uma publicação em conjunto em algum periódico também utilizado pelo pesquisador 3251. Entre a segunda e terceira linha da tabela também se verifica um empate no fator de recomendação, porém nesse caso não há publicação em conjunta entre os pesquisadores 3260 e 3270 apenas entre eles e o pesquisador 3251.

## 7. CONCLUSÃO

O artefato mostrado nesse trabalho pode assumir um papel importante no auxílio a gestão de um programa de pós-graduação, principalmente considerando a avaliação da CAPES, por sugerir aos gestores e demais pesquisadores possibilidades de realização de parcerias baseadas nas publicações já realizadas pelos pesquisadores. Tal sugestão é viabilizada por meio da recomendação de parcerias empregada como ferramenta de melhoria nos índices de publicações acadêmicas, se mostrando, uma solução viável para ser utilizada em

diversos cenários nos quais haja necessidade de melhoria dos indicadores de produção através do aumento das relações de coautoria entre os pesquisadores.

A ferramenta aqui apresentada é facilmente aplicada a outros programas de pós-graduação no país, uma vez que foi utilizado como fonte os dados presentes nos currículos Lattes dos pesquisadores vinculados aos programas. Além disso, caso sejam realizadas as devidas adaptações, diversas bases de dados podem ser utilizadas ampliando ainda mais as potencialidades de recomendação da ferramenta. A ferramenta está publicamente disponibilizada através do endereço <http://scicoll-academicmetrics.rhcloud.com/>.

Como trabalhos futuros pretende-se validar as recomendações realizadas pelo módulo, assim como expandir a base de dados utilizada de forma a abranger pesquisadores de diversos programas nacionais e internacionais. Pretende-se ainda experimentar outras abordagens no modelo de recomendação, principalmente a adoção de sistemas multiagentes em um modelo multi-camadas, no qual cada camada representa uma forma de interação entre pesquisadores como coautoria, participação em projeto, participação em banca apenas para citar algumas possibilidades.

## 8. REFERÊNCIAS

- [1] Hector N. Melo, Ruben A. Perorazio, and Jonice Oliveira. Ambiente analítico web para análise da colaboração científica no cenário médico. *X Simpósio Brasileiro de Sistemas de Informação (SBSI)*, 2014.
- [2] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. 2009.
- [3] Edeilson M. Silva, Ricardo A. Costa, Mario Godoy Neto, Robson Y. S. Oliveira, and Silvio R. L. Meira. Promovendo melhorias na comunicação e colaboração em uma plataforma de gestão de conhecimento através de recomendações. *V Simpósio Brasileiro de Sistemas de Informação (SBSI)*, 2009.
- [4] Edmilson Barcelos Rocha, Mariano Pimentel, and Morganna Carmem Diniz. Desenvolvimento de um modelo da participação em batepapo seguindo a abordagem design science research. *X Simpósio Brasileiro de Sistemas de Informação (SBSI)*, 2014.
- [5] Victor Ströele, Geraldo Zimbrão, and Jano M. Souza. Group and link analysis of multi-relational scientific social networks. *J. Syst. Softw.*, 86(7):1819–1830, July 2013.
- [6] Vinícius P. Freire and Daniel R. Figueiredo. Ranking in collaboration networks using a group based metric. *Journal of Brazilian Computer Society*, 17(4):255–266, November 2011.
- [7] Vijay K Vaishnavi, William Kuechler, and William Kuechler Jr. *Design science research methods and patterns: innovating information and communication technology*. Crc Press, Oct, 30 2007.
- [8] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O’Reilly, 2013.
- [9] Jim Webber. A programmatic introduction to neo4j. In *Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity, SPLASH ’12*, pages 217–218, New York, NY, USA, 2012. ACM.
- [10] Claudio Tesoriero. *Getting Started with OrientDB*. Packt Publishing Ltd, 2013.
- [11] Matías Javier Antiñanco. *Bases de Datos NoSQL: escalabilidad y alta disponibilidad a través de patrones de diseño*. PhD thesis, Facultad de Informática, Universidad Nacional De La Plata, 2014.
- [12] Dan Brickley and Ramanathan V Guha. Resource description framework (rdf) schema specification 1.0: W3c candidate recommendation 27. March 2000.
- [13] Ravi Kothuri, Albert Godfrind, and Euro Beinat. *Pro oracle spatial for oracle database 11g*. Dreamtech Press, 2008.
- [14] Regina Obe and Leo Hsu. *PostGIS in action*. Manning Publications Co., 2011.
- [15] Prem Melville and Vikas Sindhwani. Recommender systems. In *Encyclopedia of machine learning*, pages 829–838. Springer, 2010.
- [16] Trygve Reenskaug. Mvc. <http://heim.ifi.uio.no/~trygver/themes/mvc/mvc-index.html>, Access: Mar, 2 2015.
- [17] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.
- [18] Massimo DiPierro. *Web2Py Enterprise Web Framework, 2Nd Ed*. Wiley Publishing, 2nd edition, 2009.
- [19] V. Batagelj and A. Mrvar. Pajek-program for large network aalysis. *Connections*, 21:47–57, 1998.
- [20] Graham R Gibbs. *Qualitative data analysis: Explorations with NVivo*. Open University, 2002.
- [21] Analytic Technologies. Ucinet 6 social network analysis software. <http://www.analytech.com/ucinet.htm>, 2013.



