

Scientific Collaboration Networks of the Academic Brazilian Community of HCI

Felipe Ciacia de Mendonça
PPGCA
UDESC
Joinville, SC, BR
felipe.mendonca17@edu.udesc.br

Isabela Gasparini
PPGCA, Depto de Ciência da
Computação - UDESC
Joinville, SC, BR
isabela.gasparini@udesc.br

Rebeca Schroeder
Depto de Ciência da Computação
UDESC
Joinville, SC, BR
rebeca.schroeder@udesc.br

Milene Selbach Silveira
Escola Politécnica, PUCRS
Porto Alegre, RS, BR
milene.silveira@pucrs.br

Simone Diniz Junqueira Barbosa
Depto de Informática, PUC-Rio
Rio de Janeiro, RJ, BR
simone@inf.puc-rio.br

ABSTRACT

Social networks enable the creation of specific groups and communities such as scientific collaboration networks, which are composed of their authors, who in turn are connected through their publications. This work generated the networks of scientific collaboration of the Brazilian Community of Human-Computer Interaction analyzing the most prolific authors of the Brazilian Symposia on Human Factors in Computing Systems (IHC). First an extraction tool was developed in order to collect data from Lattes Platform. After the process of preparation, data modeling based on classification techniques, cleaning and standardization of data mining techniques were used. Finally, scientific collaboration networks were generated and analyzed through the bibliometric analysis and social network analysis techniques. The results show IHC community network in relation to its patterns and relationships. These data can help researchers get insights and increase collaboration among their peers.

CCS CONCEPTS

- Human-centered computing → Human computer interaction (HCI)
- Information systems → Information systems applications → Collaborative and social computing systems and tools

KEYWORDS

IHC, Scientific Collaboration Networks, Social Network Analysis, Data Mining, Bibliometric Analysis.

ACM Reference format:

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

IHC 2018, October 22–26, 2018, Belém, Brazil
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-6601-4/18/10...\$15.00
<https://doi.org/10.1145/3274192.3274228>

Felipe Ciacia de Mendonça, Isabela Gasparini, Rebeca Schroeder, Milene Selbach Silveira and Simone Diniz Junqueira Barbosa. Scientific Collaboration Networks of the Academic Brazilian Community of HCI. In *17th Brazilian Symposium on Human Factors in Computing Systems (IHC 2018)*, October 22–26, 2018, Belém, Brazil. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3274192.3274228>

1 Introdução

Segundo Balancieri [1], redes sociais, são um conjunto de pessoas ou grupos que possuem conexões com um ou com todos os outros indivíduos de uma mesma rede. As pessoas seriam chamadas de “atores”, e as conexões de “relacionamentos” ou “ligações”. O relacionamento entre os dois indivíduos pode ser uma amizade, uma colaboração científica ou um membro em comum entre dois grupos, entre outros tipos de relacionamentos [1]. Assim, pode-se dizer que os membros com o maior número de conexões são os pontos mais valiosos para espalhar uma notícia ou informação. A análise de redes sociais está ligada a conceitos aplicados à teoria dos grafos, e desta forma, grafos podem ser utilizados para representar as relações entre os autores. Esta análise também envolve algumas técnicas de mineração de dados [19].

Segundo Weisz e Roco [21], a colaboração científica pode ser um empreendimento cooperativo que procura desenvolver trabalhos envolvendo metas em comum, esforço coordenado e resultados ou produtos (trabalhos científicos) com responsabilidade e mérito compartilhados. Assim, os autores consideram que a colaboração científica oferece uma fonte de apoio para melhorar o resultado e maximizar o potencial da produção científica.

Porém, apesar da colaboração científica existir há muito tempo, a análise e investigação das redes de coautoria para explorar a colaboração científica entre pesquisadores constituem uma área relativamente nova [20]. O processo de disseminação de informações em redes sociais não é simples. A assimilação das informações pelos membros da rede depende de uma série de fatores, como por exemplo, interesse, percepção acerca do contexto e confiabilidade, dentre outros fatores sociológicos, psicológicos e antropológicos [6].

Segundo Gabardo [6], compreender quais são os autores mais influentes em uma rede social científica é de grande interesse a

todos os pesquisadores. Também é possível saber quais são os autores mais produtivos e influentes dentro de uma comunidade, detectar os membros quanto a sua região geográfica, as suas produções, a internacionalização de sua pesquisa, as instituições nas quais é vinculado, e as próprias redes de coautoria das quais faz parte. Compreender como se formam essas redes de colaboração científica é importante para entender as preferências, padrões, e como se dá esse relacionamento entre os autores de uma comunidade científica [17].

Este trabalho extraiu os autores mais prolíficos da comunidade brasileira de Interação Humano-Computador (IHC) por meio da base de dados do Simpósio Brasileiro sobre Fatores Humanos em Sistemas computacionais (IHC) (dados obtidos por Gasparini et al. [7]), mas estendeu a pesquisa, observando todas as publicações cadastradas na Plataforma Lattes destes autores nos diversos veículos de publicação, suas orientações de mestrado e doutorado e quais instituições pertencem. O trabalho visa gerar as redes de colaboração científica da comunidade brasileira de IHC, e, para isso, foi criado um modelo para transformação dos dados em conhecimento por meio de técnicas de mineração de dados.

Os dados foram inicialmente capturados dos Anais do Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais, selecionando os autores mais prolíficos que, posteriormente, foram pesquisados na Plataforma Lattes para obtenção dos seus currículos. A partir disso, foi implementada a extração automática dos dados dos currículos selecionados e, em seguida, foi efetuada a padronização e limpeza das informações. Os dados resultantes foram armazenados em um banco de dados orientado a grafos, a partir do qual foram geradas as redes de colaboração científica e realizadas as análises da comunidade brasileira de IHC.

Este trabalho está estruturado como segue. Na seção 2 são apresentados os conceitos fundamentais sobre Análise de Redes Sociais, explorando as redes de colaboração científicas e as métricas utilizadas. Na seção 3 são apresentados os trabalhos relacionados. Na seção 4 é detalhado o processo de desenvolvimento do trabalho. Por fim, as análises dos resultados são apresentadas e, em seguida, a conclusão do trabalho.

2 Análise de Redes Sociais

A área de pesquisa denominada Análise de Redes Sociais (SNA - *Social Network Analysis*) estuda, por meio da identificação dos atores e suas ligações, as relações entre eles de forma a poder identificar suas formas de interação, contribuindo para o conhecimento sobre a rede social e o seu desenvolvimento. SNA está focada na estrutura dos relacionamentos dos indivíduos e pode abstrair e representar um fenômeno de interação e relacionamento entre unidades.

Uma Rede Social é uma estrutura social composta por indivíduos ou organizações na qual os vértices, também chamados de nós (ou vértices), são conectados por um ou mais tipos de relacionamentos de interdependência, como, por exemplo, amizade, trabalho, crença [5]. Dessa forma, a estrutura de dados por meio de grafos é uma das mais adequadas para realizar a representação computacional das redes sociais [5]. A rede social é modelada de forma que os nós podem representar os atores e as arestas a relação entre estes autores, formando diversas características, entre elas, a centralidade, que se refere à posição de um nó na

estrutura de um grafo, mostrando assim a importância e a relação do nó com os outros indivíduos [3].

O crescimento da Web permitiu diferentes ferramentas para a interação entre pessoas, tais como diferentes redes sociais. Um tipo de rede social são as redes de colaboração científicas ou acadêmicas. Nestas redes os vértices representam os pesquisadores, e as arestas as colaborações científicas [11].

2.1 Redes de Colaboração Científica

Quando há um conjunto de pessoas ou grupos que possuem conexões de algum tipo com um ou mais integrantes de uma rede, esta rede pode ser considerada uma rede de colaboração na qual o grupo busca trabalhar de forma cooperativa, procurando o desenvolvimento de esforços e desenvolvimento coordenado a fim de atingir metas em comum [14],[21]. Entre as redes de colaboração existentes, as que são formadas no âmbito científico são as Redes de Colaboração Científica formada por acadêmicos, professores e pesquisadores.

Tal colaboração pode ser formada desde uma orientação ou discussão de ideias até mesmo uma participação ativa em um projeto de pesquisa específico. O pesquisador pode ser considerado um colaborador ou aparecer como coautor por ter fornecido um material ou auxiliado no trabalho de alguma forma [1]. A coautoria na pesquisa científica auxilia no intercâmbio, nas relações entre autores em uma determinada área do conhecimento e nas experiências interdisciplinares [17].

2.2 Métricas utilizadas em Análises de Redes Sociais

Uma das análises mais interessantes da área de SNA é determinar quais são os nós mais importantes [9]. Para isso, há diversas métricas para serem usadas; porém, a aplicação delas para questões específicas exigem o bom senso e o foco da análise daqueles que estiverem realizando a pesquisa com as redes sociais para poderem ter uma compreensão adequada da rede [16].

A Centralidade é o termo usado para descrever o quão importante é um nó dentro de uma rede, podendo ser calculada de várias formas, entre as quais foram escolhidas para este artigo os principais métodos amplamente usados para identificar este nó influente (vértice) dentro de uma rede social (grafo) que são: Centralidade de Grau, Centralidade de Proximidade e a Centralidade de Intermediação [9], [10]. Além disso, foram escolhidas tais métricas pois o banco de dados Neo4j suportava a extração destas métricas da rede formada.

2.2.1 Centralidade de Grau. A centralidade de Grau trata da contagem do número de conexões diretas que um vértice possui, isto pode ser visto como a probabilidade que o vértice tem de receber ou contribuir com alguma informação dentro da rede. Dentre as três métricas citadas, a centralidade de grau é a mais elementar, por determinar a importância de um vértice exclusivamente pela quantidade de vértices adjacentes a ele.

2.2.2 Centralidade de Proximidade. Essa métrica mede a proximidade de um nó em relação ao grafo inteiro baseado na soma do inverso das distâncias do vértice escolhido aos demais vértices do grafo. Assim sendo, esta medida só pode ser calculada para grafos conexos. Esta medida pode ser interpretada como uma medida de rapidez, já que podemos determinar a “velocidade” que

algum nó precisará para difundir uma informação entre os outros nós da rede. Logo, seja D uma matriz simétrica cujo elemento d_{ij} representa a menor distância do vértice i para o j , tal que $i \geq 1, j \leq n$, e n é o número total de vértices do grafo. O cálculo da centralidade de proximidade (CP) do vértice i é dado por:

$$CP(i) = \frac{1}{\sum_{j=1}^n d_{ij}}$$

De acordo com a definição, o vértice mais central do grafo é aquele com o maior valor de CP.

2.2.3 Centralidade de Intermediação. A centralidade de intermediação atribui importância a um vértice da rede em função da passagem de fluxo por ele para interligar outros dois vértices através do menor caminho entre eles [9]. A centralidade de intermediação (CI) de um vértice i é dada por:

$$CI(i) = \sum_{j < k} \frac{C_{jik}}{C_{jk}}, i \neq j, 0 < j < k \leq n$$

Na definição, C_{jk} representa a quantidade de caminhos mínimos entre os nós j e k , e C_{jik} representa a quantidade de caminhos mínimos entre os nós j e k que passam por i . O total de vértices da rede é representado por n . Logo, o vértice com maior centralidade de intermediação é aquele que participa de maneira mais ativa em um processo de interação, onde os caminhos mínimos são percorridos. Portanto, essa métrica quantifica o número de vezes que um vértice serviu de ponte ao longo do caminho mais curto entre outros dois vértices. Seria como um nó que interliga outros vértices independentes ou até conjuntos de nós independentes, e que serve como ponte para o fluxo de informações entre duas redes distintas.

3 Trabalhos Relacionados

Diversos trabalhos na literatura envolvem análises de redes sociais extraídas de cooperações de produção científica. Com foco na comunidade científica brasileira, a grande maioria dos trabalhos se baseia na extração de dados da Plataforma Lattes, com o complemento de outros tipos de fontes de dados em alguns casos. Algumas ferramentas se inserem neste contexto, como scriptLattes e a plataforma Sucupira. No caso do scriptLattes [13], a ferramenta recebe currículos Lattes de um grupo de interesse, compila as listas de produções científicas e orientações acadêmicas para formar relatórios e grafos de coautoria entre os membros do grupo. Em relação à Sucupira, dentre suas principais funcionalidades está a análise e a comparação de desempenho de pesquisadores, bem como a visualização de redes sociais acadêmicas entre eles.

Além da Plataforma Lattes, um conjunto de trabalhos utiliza anais de conferências para extrair coautorias e efetuar algumas análises. O trabalho de Silva et al. [17] realiza uma pesquisa exploratória dos anais do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) para formar redes de coautoria. Para enriquecer os dados, algumas informações dos autores relacionados são extraídas da Plataforma Lattes. Técnicas de análise de redes sociais são utilizadas para identificar os autores mais produtivos, colaborações entre instituições e produtividade

por região geográfica. Um trabalho similar é proposto por Steinmacher et al. [18], neste caso para a comunidade do Simpósio Brasileiro de Sistemas Colaborativos (SBSC). Neste trabalho os autores conseguiram demonstrar a expansão dos grupos de pesquisa que deram origem a comunidade. Em Bartneck e Hu [2], a rede de coautoria de artigos publicados na conferência *Conference on Human Factors in Computing Systems* é analisada. Neste caso constatou-se que a diversidade de organizações envolvidas em um artigo não estava associada a uma maior citação do artigo, mas sim a um maior número de premiações.

Digiampretri e Silva [5] desenvolveram um *framework* para realizar análises nas redes sociais de pesquisadores extraídas da Plataforma Lattes. Algumas técnicas como Análise de Redes Sociais, Extração de Conhecimento e Teoria dos Grafos foram combinadas para a extração de métricas como medidas de centralidade, bem como a seleção, agrupamento e classificação de dados. O foco do trabalho era produzir indicadores para avaliar grupos de pesquisadores de programas de pós-graduação. O mesmo grupo de autores publicou uma série de técnicas de extração de dados a serem utilizadas sobre a Plataforma Lattes para produzir um banco de dados de mais de um milhão de currículos Lattes [4]. Além disto, o referido trabalho apresentou *insights* para a aplicação de técnicas de mineração de dados sobre o banco gerado. Em Maruyama e Digiampretri [11] os trabalhos anteriores são estendidos com técnicas de predição de relacionamentos em redes de colaboração científica.

Este artigo tem por objetivo gerar redes de colaboração científica da comunidade brasileira de IHC. Assim como no trabalho de Digiampretri e Silva [5], técnicas de mineração de dados são aplicadas passando por diversas fases como pesquisa, compreensão, preparação, modelagem e avaliação de dados. Estas fases e demais detalhamentos do desenvolvimento da pesquisa serão descritas nas seções a seguir.

4 Desenvolvimento

Este trabalho seguiu um modelo de extração do conhecimento desenvolvido com base nas técnicas de KDD (*Knowledge Discovery in Databases*), e está dividido em três diferentes etapas. Na primeira foram selecionados os autores mais prolíficos de IHC, para então extrair seus respectivos currículos da Plataforma Lattes. Na segunda etapa foram interpretados os dados para serem utilizados, preparando-os para as etapas posteriores, aplicando transformações nos dados, quando necessário. Na terceira etapa, com os dados devidamente prontos, foram geradas as redes de colaboração científica da comunidade de IHC. Esse processo é descrito abaixo:

- 1) Seleção dos Autores mais Prolíficos de IHC
- 2) Download dos Currículos em XML da Plataforma Lattes
- 3) Implementação do Script em PHP para Extração Automática de Informações
- 4) Inserção no banco de dados relacional MySQL
- 5) Padronização, Limpeza e Filtragem dos dados
 - a) Transformação de dados através de scripts em Python
 - b) Inserção no banco de dados Neo4j
- 6) Análises de Redes Sociais
- 7) Análises Estatísticas e Bibliométricas

4.1 Modelo de Extração KDD

O modelo desenvolvido abrange a fase de Seleção e Extração de Dados que capturou os currículos da Plataforma Lattes dos autores selecionados conforme a Figura 1. Em seguida, houve a implementação de um *script* para extrair as informações contidas nos currículos e inserir em um banco de dados relacional. Já na etapa de Interpretação e Preparação dos Dados, foram realizados a Filtragem, Limpeza e Padronização dos Dados. Ainda nesta etapa, com os dados padronizados, algumas tabelas de dados precisaram ser adaptadas por meio de *scripts* de Transformação de Dados para ficarem no formato adequado para a inserção no banco de dados orientado a grafos, para posteriormente realizar a Geração das Redes. O desenvolvimento de todas as etapas do modelo durou 12 meses, de setembro de 2016 a setembro de 2017.

4.1.1 Seleção e Extração dos Dados. Os autores selecionados para esta pesquisa foram os mais prolíficos da comunidade de IHC no Brasil, usando como base os anais do Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (Simpósio IHC). Como critério de filtragem foram utilizados os autores que tiveram cinco ou mais artigos completos no Simpósio IHC, considerando as edições entre 1998 até 2015. Como resultado há uma seleção de 29 autores. Esse resultado corrobora com o trabalho de Gasparini et al. [8]. A Figura 1 apresenta esses 29 autores em ordem alfabética.

Artur Henrique Kronbauer	Marcelo Soares Pimenta
Carla Faria Leitão	Marco Antônio Alba Winckler
Carla Maria Dal Sasso Freitas	Maria Cecília Calani Baranauskas
Celso Alberto Saibel Santos	Maria Elizabeth Sucupira Furtado
Clarisse Sieckenius de Souza	Milene Selbach Silveira
Cristiano Maciel	Raquel Oliveira Prates
Denis Silva da Silveira	Roberto Pereira
Heloísa Vieira da Rocha	Sérgio Roberto Pereira da Silva
Isabela Gasparini	Simone Bacellar Leal Ferreira
Jair Cavalcanti Leite	Simone Diniz Junqueira Barbosa
Janne Y. Y. Oeiras Lachi	Tayana Uchôa Conte
Junia Coutinho Anacleto	Vania Paula de Almeida Neris
Lara S. Godoy Piccolo	Vinícius Carvalho Pereira
Leonardo Cunha de Miranda	Walter de Abreu Cybis
Lucia Vilela Leite Filgueiras	

Figura 1. Autores mais prolíficos do IHC

Após ter o conhecimento de quais autores iriam ser utilizados como base da pesquisa, foram extraídos manualmente da Plataforma Lattes os 29 currículos destes autores, no formato *XML* (*Extensible Markup Language*). Esta etapa foi feita de forma manual, no período do mês de setembro de 2016. Os currículos em *XML* são compostos por diversas *tags*, estruturando o currículo em diversas categorias, sendo as principais:

- **Dados Gerais:** Informações de dados pessoais, resumo do currículo *vitae*, endereço, formação acadêmica, atuações profissionais, área de atuação e idiomas;
- **Produções Bibliográficas:** Informações de trabalhos em eventos, artigos publicados, livros e capítulos de livros publicados, textos publicados em jornais ou revistas e demais produções bibliográficas;
- **Produções Técnicas:** Informações sobre produtos de softwares desenvolvidos, trabalhos técnicos produzidos e demais tipos de produções técnicas;
- **Outras Produções:** Informações das orientações concluídas e demais trabalhos realizados;

- **Dados Complementares:** Informações das participações em banca, participações em eventos e congressos, orientações em andamento e outras informações adicionais sobre instituições e cursos.

Utilizando uma extensão da API chamada *SimpleXML*, foi implementado um *script* em *PHP* para leitura dos arquivos *XML* e armazenamento das informações em tabelas no banco de dados *MySQL*.

4.1.2 Interpretação e Preparação dos Dados. Os dados foram separados em tabelas para manter as informações organizadas, bem como para facilitar a filtragem, padronização e limpeza dos dados para que as análises pudessem ser feitas. Além disso, para a inserção no banco de dados orientado a grafos *Neo4j*, as tabelas devem estar em um formato padronizado. Desta forma, foi necessária uma conversão de algumas tabelas, para que pudessem ser inseridas no banco *Neo4j*, e então gerar as redes de colaboração científicas.

A limpeza e padronização dos dados consistiu na correção de informações faltantes, errôneas ou inconsistentes das tabelas, e além disso, a padronização de todos os nomes dos autores, nomes das instituições e nomes dos veículos de publicação. Esta etapa foi um dos grandes problemas enfrentados, pois foram encontrados vários problemas de padronização de nomes, como nomes diferentes para mesmos eventos, pessoas e instituições. Um exemplo foi o próprio Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais. Foram encontrados mais de 20 nomes diferentes para o evento. Isso se deve a mudanças no nome do evento ao longo do tempo, erro de escrita de autores, ou até mesmo por causa da edição do evento que acompanhou o nome. Este processo de padronização e limpeza foi iniciado em maio de 2017 e finalizado em meados de junho do mesmo ano.

Após a fase de interpretação e pré-processamento dos dados, em que os dados foram limpos, padronizados e filtrados, as tabelas já estavam aptas para realizar análises bibliométricas e estatísticas. No entanto, algumas dessas tabelas ainda não estavam no formato adequado para inserção no banco de dados *Neo4j*, porque alguns requisitos são necessários para a criação de relacionamentos no banco. O *Neo4j* é um banco de dados orientado a grafos no qual é possível criar nós equivalentes às entidades de um banco de dados relacional e as relações entre esses nós, que seriam as relações de um banco de dados relacional convencional. Porém, o diferencial do banco de dados *Neo4j* é a sua representação na forma de grafos, o que torna o processamento muito mais rápido em virtude da representação de redes sociais se adequarem ao modelo de grafos, além de gerar a representação visual das redes sociais formadas e também permitir a extração de métricas tais como as centralidades. Desta forma, a organização dos dados se tornou mais simples, não precisando de várias tabelas como no banco de dados relacional.

Desta forma foi necessário implementar *scripts* em *Python*, que recebiam como entrada as tabelas de dados geradas pelo banco *MySQL* e retornavam as mesmas tabelas, porém com os requisitos necessários para inserção no banco de dados em grafos *Neo4j*. Com as tabelas devidamente prontas, foi possível gerar as redes de colaboração científica.

4.1.3 Geração das Redes. A Figura 2 representa o modelo de dados no banco *Neo4j*. Um grafo registra nós e relacionamentos, de modo que os relacionamentos representam associações entre

nós. O modelo utilizado pelo *Neo4j* estabelece a possibilidade de existirem propriedades para nós e relacionamentos. Para a criação do banco de dados em grafo, os registros de tabelas de entidade do banco relacional foram mapeados como nós, e seus campos como propriedades destes nós. Os relacionamentos entre nós foram obtidos através de registros de tabelas associativas ou chaves estrangeiras do relacional. O resultado deste mapeamento produziu um grafo que representa a rede de colaboração científica da comunidade IHC.

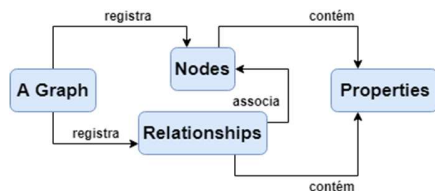


Figura 2. Representação do Modelo de Dados do Neo4j¹

Os nós inseridos foram divididos em cinco tipos, com suas respectivas descrições e quantidades inseridas no banco *Neo4j*. A Tabela 1 descreve estes tipos de nó.

Tipo de Nó	Descrição	Quantidade
Pesquisador	Autores mais prolíficos de IHC	29
Orientado	Orientados/Coorientados de Mestrado e/ou Doutorado	624
Coautor	Coautores dos pesquisadores	2316
Veículo	Veículos de Publicação	1199
Instituição	Universidades as quais os autores são vinculados	43

Tabela 1. Tipos de nós

Após a inserção dos nós, foram criadas conexões entre os nós, que também foram divididas em cinco tipos, conforme é apresentado na Tabela 2.

Tipo de Relação	Descrição	Quantidade
Doutorado	Orientações/Coorientações de Doutorado	153
Mestrado	Orientações/Coorientações de Mestrado	547
Coautoria	Coautoria entre os autores das publicações	2949
Publicou	Conexão entre os pesquisadores e os veículos que ele publicou	2093
Vínculo	Vínculo entre os pesquisadores e a Universidade a qual ele é vinculado	659

Tabela 2. Conexões entre os nós

Portanto foram criados cinco tipos de nós no banco de dados em grafos, com cinco tipos de relacionamentos entre eles, formando assim, uma Rede de Colaboração Científica da Comunidade Brasileira de IHC. O grafo que representa esta geração está apresentado na Figura 3, que ilustra os tipos de nós e de relacionamentos gerados no *Neo4j*.

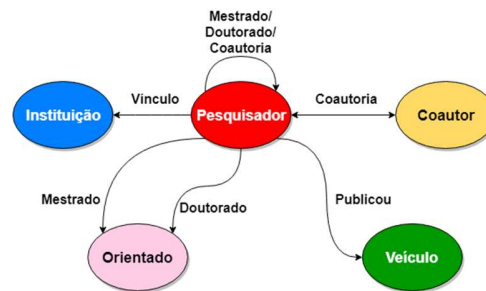


Figura 3. Representação geral da Rede gerada

Vale ressaltar que a rede formada possui algumas restrições. A primeira é que ela é estática (não é uma rede temporal) e que cada relação é representada por uma única aresta na rede. Desta forma um autor prolífico ao orientar uma pessoa em seu mestrado e doutorado e publicar vários artigos em conjunto, gera na rede de orientação duas arestas distintas entre eles (uma na cor amarela e outra na cor azul) e uma outra aresta de coautoria. Da mesma forma se um pesquisador prolífico orientou outro prolífico isso fica representado na rede por meio de dois nós em vermelho (prolíficos). Exemplos destas representações são apresentados na próxima seção.

5 Análise dos Resultados

Com base nos dados extraídos dos currículos dos autores mais prolíficos do IHC e das redes geradas, foi possível observar diversos aspectos das redes de colaboração dos pesquisadores envolvidos. Como exemplo, apresentamos as redes formadas por relações de orientação de Mestrado e Doutorado, bem como as Universidades relacionadas aos respectivos pesquisadores. Além destas, serão abordadas redes formadas por relacionamentos de coautoria e os veículos de publicação associados. Lembrando que, para a construção destas redes não foi considerado a questão de temporalidade das publicações e orientações, mas somente a quantidade de trabalhos contribuídos ao longo dos anos. Algumas métricas foram obtidas a partir da análise das redes sociais formadas como, por exemplo, medidas de centralidade que determinam a importância de um nó dentro do grafo. No caso das redes geradas neste trabalho, foi possível identificar os nós (Pesquisadores) centrais ou mais influentes dentro desta rede social. Além das centralidades, também foi possível verificar a formação de grupos ou comunidades, com o objetivo de definir conjuntos de nós que possuem mais conexões entre si do que com o resto da rede social.

5.1 Visão Completa

A Rede de Colaboração Científica da Comunidade Brasileira de IHC completa e conexa, como pode ser observada na Figura 4, foi construída a partir de algumas subredes, a saber: Relacionamentos de Coautoria entre os pesquisadores prolíficos e seus coautores; Relacionamentos entre os pesquisadores prolíficos e seus orientados de Mestrado ou Doutorado, além das respectivas ligações com suas Universidades tanto para os pesquisadores

¹ <http://neo4j.com/developer/graph-database/>

quanto para os orientados e; Relacionamentos entre os pesquisadores prolíficos e os veículos nos quais ele publicou. Para ficar clara a distinção dos nós na rede, foram atribuídas cores para cada tipo de nó especificado na Tabela 3.

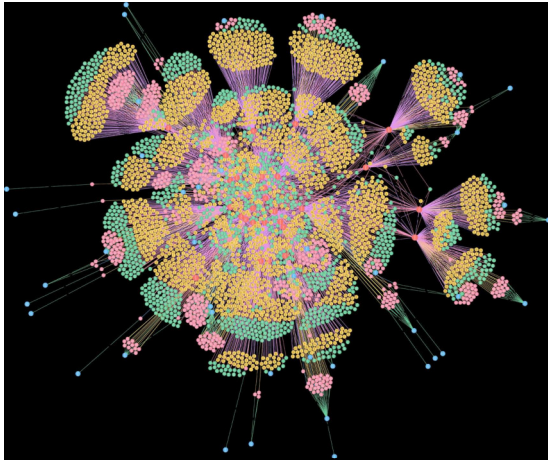


Figura 4. Rede de Colaboração Científica Completa da Comunidade Brasileira de IHC

Cor	Tipo de Nó
Azul	Instituição/Universidade
Verde	Veículo de Publicação
Vermelho	Pesquisador Prolífico
Amarelo	Coautor
Rosa	Orientados de Mestrado e Doutorado

Tabela 3. Legenda de Cores para Redes de Colaboração

5.2 Rede de Orientados de Mestrado e Doutorado

A subrede formada com os relacionamentos entre os pesquisadores e seus orientados de Mestrado e Doutorado é mostrada na Figura 5.

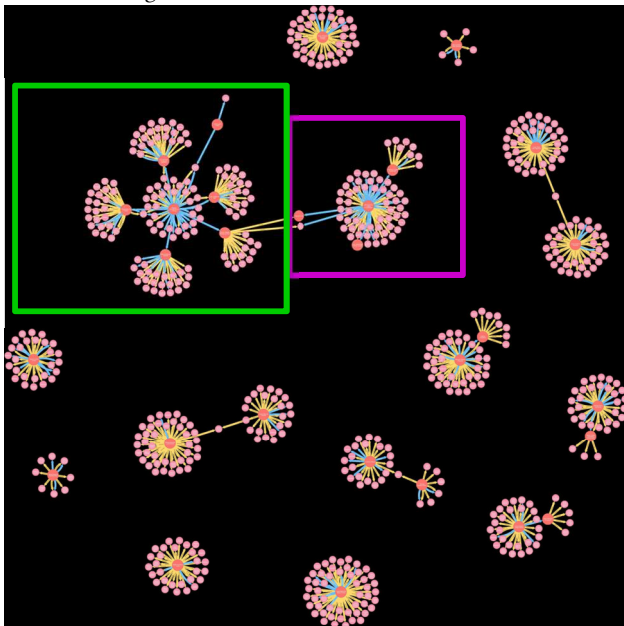


Figura 5. Rede de orientação (mestrado e doutorado)

É possível visualizar que existem relações de orientações de Mestrado e de Doutorado entre os próprios pesquisadores prolíficos (as relações de Mestrado foram coloridas de Amarelo, enquanto as relações de Doutorado foram coloridas com a cor Azul). Observa-se que diversos pares de pesquisadores e orientados são formados, inclusive com alguns pesquisadores com orientados em comum, além de uma grande ilha formada por 11 pesquisadores com o nó central representado por uma das pesquisadoras mais influentes e sêniores da área de IHC, a professora Clarisse Sieckenius de Souza, que orientou diversos pesquisadores de Doutorado que se tornaram autores prolíficos dentro da comunidade de IHC (destacado em verde na Figura). Outra grande rede apresentada tem nó central a pesquisadora M. Cecília C. Baranauskas (destaque em roxo na Figura).

Ao fazer um recorte da Figura 5 quanto ao destaque em verde, pode-se observar com detalhes a subrede formada por pesquisadores que foram orientados/coorientados pela professora Clarisse Sieckenius de Souza, conforme Figura 6. Observa-se que ela orientou diversos autores prolíficos (Jair C. Leite, Milene S. Silveira, Sérgio Roberto P. da Silva, Simone D. J. Barbosa e Raquel O. Prates). As arestas em amarelo representam orientação de mestrado e as em azul a orientação de doutorado. Também é possível verificar que a professora Clarisse teve coorientação juntamente com alguns destes autores prolíficos. Já o recorte em roxo da Figura 5 apresenta a subrede de orientação formada como ponto central a professora M. Cecília C. Baranauskas, conforme Figura 7.

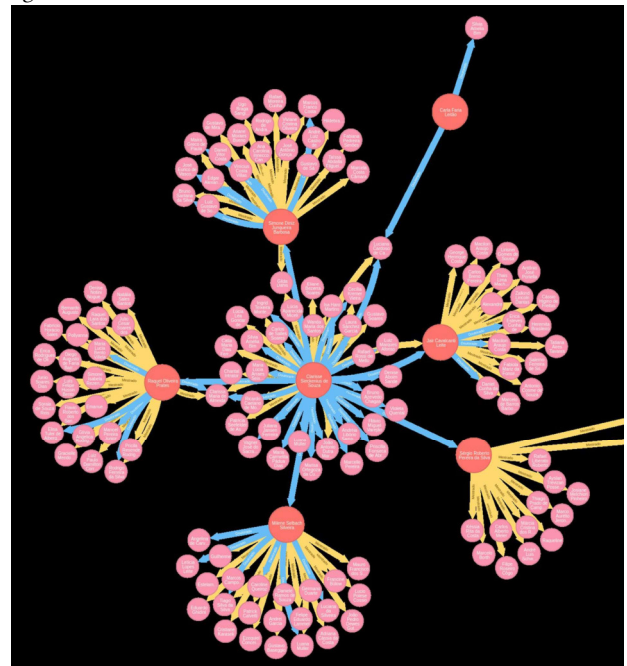


Figura 6. Subrede de orientação - Clarisse Sieckenius de Souza

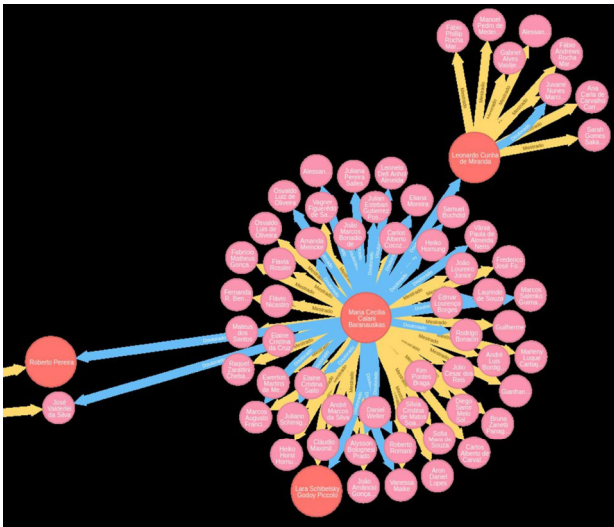


Figura 7. Subrede de orientação - M. Cecília C. Baranauskas

5.3 Rede de Orientados, Pesquisadores e suas respectivas Universidades

A Figura 8 apresenta a rede de orientação e instituições. Os nós em azul representam instituições e os nós na cor rosa pesquisadores orientados. Como há poucas ligações entre instituições via um orientado é possível ver que existem poucos orientados que fizeram o Mestrado e o Doutorado em Instituições distintas. Além disso existem poucas Universidades com mais de um Pesquisador Prolífico em IHC. Entre as Universidades que mais se destacam pela quantidade de orientados estão a Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), a Universidade Federal do Rio Grande do Sul (UFRGS), e a Universidade Estadual de Campinas (UNICAMP).

5.4 Rede de Pesquisadores e os veículos de publicação

Na Figura 9 é possível visualizar a grande quantidade de veículos de publicação que foram encontrados nos currículos dos 29 autores mais prolíficos. Ao total foram encontrados 1199 Veículos de Publicações distintos.

5.5 Redes de Coautoria dos pesquisadores

A Rede de Coautoria é a principal subrede desta pesquisa e também o grafo com maior número de elementos entre os pesquisadores e os autores que realizaram parceria em suas publicações, como mostrado na Figura 10. Pode-se visualizar que existem muitos autores deslocados, os quais publicaram somente em parceria com um dos autores mais prolíficos; porém há um conjunto de autores na parte central do grafo que fizeram parceria com mais de um autor prolífico.

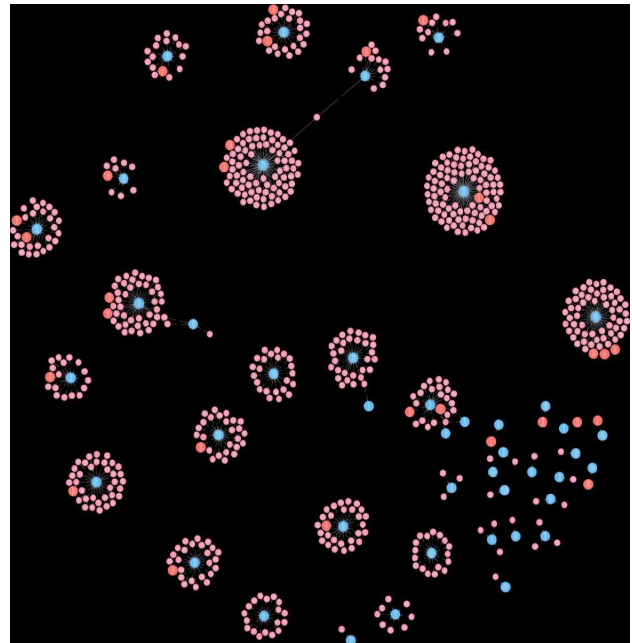


Figura 8. Rede de orientação e instituições

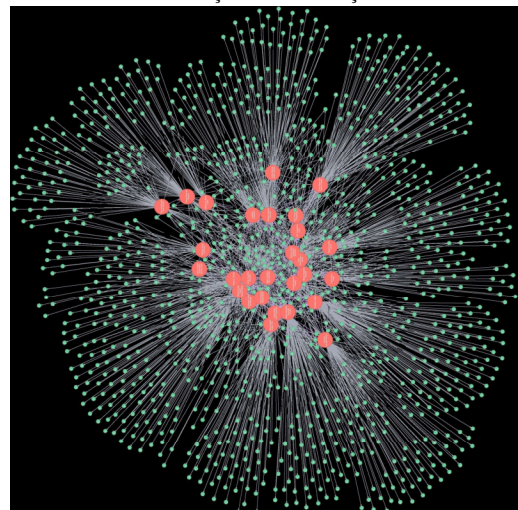


Figura 9. Rede de Pesquisadores e os veículos de publicação

Buscando investigar um pouco mais esta questão foi realizada uma consulta no *Neo4j* para que pudesse ser visto quais seriam os autores que têm maior número de conexões, i.e. que são coautores do maior número de autores prolíficos, como pode ser visto na Tabela 4. Vale ressaltar que a Tabela 4 não leva em consideração a quantidade de artigos em conjunto e sim a variedade de coautores prolíficos. Por exemplo, percebe-se que tanto o pesquisador Paulo Melo quanto o Tuomo Kujala tiveram somente um artigo em conjunto com estes oito autores. Na Figura 11 é mostrada a rede do autor Tuomo Kujala.

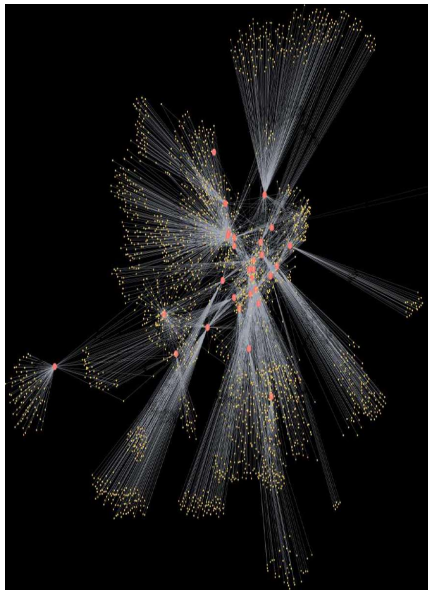


Figura 10. Redes de Coautoria dos pesquisadores

#	Coautor	Nº. coautores distintos
1	Clodis Boscarioli	8
	Heiko Horst Hornung	
	Paulo Melo	
	Tuomo Kujala	
2	Elton José da Silva	7
3	Adriana Holtz Betiol	6
	Alberto Barbosa Raposo	
	Maria Cecília Martins	
	Philippe Palanque	

Tabela 4. "Top 3" coautores em número de coautores distintos

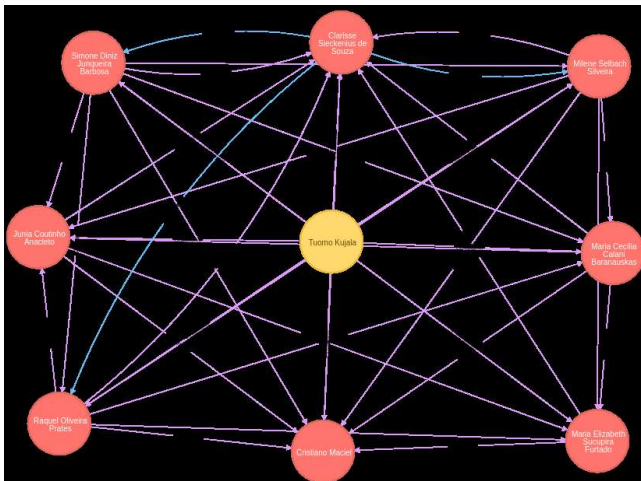


Figura 11. Rede de Tuomo Kujala – um dos autores com maior número de coautorias distintas

Os 8 pesquisadores prolíficos ligados a um dos autores com maior número de coautorias distintas, o autor Tuomo Kujala, não só tem essa ligação em comum, como são altamente interligados entre todos em forma de coautoria, além de existir algumas orientações de Doutorado e Mestrado na rede. É interessante ver a comparação

entre a Tabela 4 e a Tabela 5, em que a Tabela 5 apresenta o autor mais prolífico dentre os coautores, com base no número de artigos em conjunto, enquanto a Tabela 4 apresentou o autor com maior número de coautores distintos. Em relação a Tabela 5, observa-se que o coautor mais prolífico José Palazzo Moreira de Oliveira com 61 publicações em coautoria com os pesquisadores prolíficos não aparece no Top 3 em número de coautorias distintas já que ele publicou em parceria com apenas 3 pesquisadores, enquanto que os dois segundos lugares da Tabela 5, Heiko Horst Hornung e Philippe Palanque, com 58 publicações cada um, aparecem na Tabela 4 com 8 e 6 coautorias distintas, respectivamente.

#	Coautor	Qtde
1	José Palazzo Moreira de Oliveira	61
2	Heiko Horst Hornung	58
	Philippe Palanque	
3	Aparecido Fabiano Pinatti de Carvalho	55
	Luciana Porcher Nedel	
4	Rodrigo Bonacin	52
5	Aline da Silva Alves	51
	Avanilde Kemczinski	
6	Elaine Cristina Saito Hayashi	43
7	Marcos Alexandre Rose Silva	42
8	Maria Cecília Martins	38
9	Ana Cristina Bicharra Garcia	35
10	Katia Morosov Alonso	34

Tabela 5. Coautores mais prolífico dentre os coautores

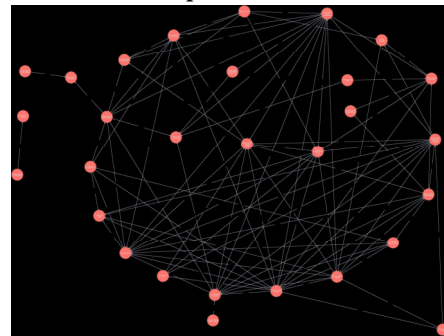


Figura 12. Rede de coautoria entre os 29 autores mais prolíficos do IHC

Ainda com base na Rede de Coautoria dos Pesquisadores Prolíficos pode-se observar a coautoria entre eles, removendo os autores que não sejam os mais prolíficos, i.e., apresentando apenas os 29 pesquisadores principais como pode ser visualizado na Figura 12. Verifica-se que os 29 pesquisadores exercem uma colaboração científica forte entre si.

5.6 Métricas de Centralidade

Com base das redes geradas, foram calculadas as métricas de Centralidade de Grau, Centralidade de Intermediação e Centralidade de Proximidade. As Tabelas 6-9 apresentam estas métricas para os pesquisadores que obtiveram as 10 melhores pontuações em cada uma.

A Tabela 6 apresenta as métricas para a Rede de Coautoria dos Pesquisadores Prolíficos. Percebe-se que os autores que se destacaram nas três centralidades foram os autores Cristiano Maciel e Raquel Oliveira Prates, enquanto outros autores, como M.

Cecília. C. Baranauskas e Marcelo Soares Pimenta, divergem em algumas centralidades. Estes achados indicam que as associações de coautoria destes pesquisadores os conectam com maior proximidade a um maior número de pesquisadores da rede, se comparado a outros pesquisadores com *scores* inferiores. A Centralidade de Grau indica os autores que tem o maior número de coautorias diretas com outros pesquisadores. Já a Centralidade de Proximidade destaca quais pesquisadores são nós centrais dentro da rede formada, pois são nós que contém uma soma do inverso das distâncias de cada nó ao restante dos nós, ou seja, podemos interpretar como os nós que são mais próximos da maioria dos pesquisadores. Enquanto que a métrica de Centralidade de Intermediação quantifica os autores em relação ao controle do pesquisador sobre a comunicação entre os outros pesquisadores dentro da rede, interpretado assim como uma ponte dentro da rede social.

Na Tabela 7 são apresentadas as métricas de centralidade para a Rede de Orientações de Doutorado. Nota-se que os autores mais influentes seguindo a Rede de Orientações de Mestrado e Doutorado foram as pesquisadoras Maria C. C. Baranauskas e a Clarisse Sieckenius de Souza, duas pioneiras da área de IHC no Brasil.

Além de medir as centralidades dos autores mais prolíficos para descobrir quais seriam os mais influentes segundo as suas produções como orientações de mestrado e doutorado, e também com suas coautorias, foram medidas as centralidades de grau e de intermediação para as Universidades. Na Tabela 8 são apresentadas as Universidades que se mostraram centrais dentro da Rede de Colaboração Científica conforme as medidas de centralidade. Por fim, foi medida a centralidade dos autores considerando toda a Rede de Colaboração Científica, ou seja, a centralidade dos autores considerando suas orientações, coautorias, universidades e veículos de publicações, como pode ser visto na Tabela 9 demonstrando os autores que são mais influentes ou que contribuíram ao longo dos anos para a comunidade de IHC no Brasil como um todo. Levando em conta os resultados, existem muitos autores influentes, mas o nome de destaque nas centralidades foi a autora Maria Cecília Calani Baranauskas.

Quanto à centralidade de grau, observa-se algumas divergências entre as métricas. Por exemplo, na Tabela 7 as pesquisadoras Tayana Uchôa Conte e Lucia Vilela L. Filgueiras aparecem apenas na centralidade de grau. Esta ocorrência pode ser explicada pela quantidade de conexões diretas de coautoria que estas pesquisadoras apresentam. Entretanto, ao observar suas conexões entre os pesquisadores da comunidade completa, a conectividade de ambas não se mostra tão alta comparada aos pesquisadores apresentados na centralidade de proximidade e intermediação. Um comportamento inverso pode ser observado para a pesquisadora Maria Elizabeth S. Furtado que apenas não aparece na centralidade de grau.

Ainda sobre a rede de coautoria, outra observação importante é o fato da pesquisadora Clarisse Sieckenius de Souza ter sido relacionada a centralidade de proximidade, mas não na centralidade por intermediação. Neste caso, verificou-se que sobre a rede de coautoria é possível identificar grupos de pesquisadores, formando *clusters* no grafo. Neste contexto, a centralidade por intermediação privilegia pesquisadores que estão conectados ao

maior número de *clusters*. Estes *clusters* podem ser representados por pesquisadores de áreas específicas. Desta forma, os pesquisadores que detêm um alto *score* para a proximidade de intermediação podem representar aqueles que estão conectados a um maior número de áreas distintas. Este achado referente à pesquisadora Clarisse Sieckenius de Souza diz respeito apenas à rede de coautorias, entretanto, é possível observar que ela está na primeira posição nas redes de orientação de mestrado e doutorado. Esse dado é apoiado pelo trabalho de Gasparini et al. [7], que destaca a professora Clarisse como uma das pioneiras na formação de pesquisadores da comunidade. Seu nome também aparece entre os 10 pesquisadores que obtiveram o melhor *score* para a centralidade de intermediação, considerando a rede de colaboração completa.

6 Conclusão

A análise de redes de colaboração científica permite conhecer uma comunidade de pesquisadores e como estes colaboram entre si. Neste artigo foram geradas redes de colaboração científica a partir dos 29 pesquisadores mais prolíficos da comunidade de IHC. As relações da rede representam colaborações de coautorias, veículos de publicação, orientações de mestrado, doutorado e instituições de vínculo. Por meio de medidas individuais de proximidade podemos combiná-las para definir medidas globais, caracterizando assim a coesão da rede completa [22]. Portanto, através das subredes formadas pelas relações de coautoria foi possível constatar a forte coesão da comunidade de IHC ao se identificar inúmeras colaborações entre seus autores.

Ao considerar o grafo formado pelas redes de colaboração, foi possível aplicar métricas de centralidade que identificaram a importância dos nós no grafo. Para os pesquisadores que obtiveram os melhores *scores* de centralidade, constatou-se que suas associações os conectam a maior parte de pesquisadores presente na rede. Efetivamente estes pesquisadores identificados como centrais acabam por impactar a rede de colaboração como um todo. A alta coesão da comunidade de IHC é certamente devida a influência destes pesquisadores centrais, bem como de seus pares.

Alguns pontos merecem ser investigados no futuro. Por exemplo, identificar o impacto da quantidade de coautores por publicação na influência exercida pelos pesquisadores da comunidade, além de olhar alguns conceitos como o de redes *two-mode-networks* que analisam redes que ligam entidades diferentes dentro de um mesmo grafo. Neste trabalho a quantidade de publicações entre dois pesquisadores não foi considerada na formação das redes. Entretanto, considera-se que as arestas do grafo de colaboração possam ser ponderadas de forma a considerar este quantitativo. Outro ponto a ser considerado está relacionado a atualização da base de dados, que neste trabalho compreendeu até o ano de 2015. Uma outra direção para este trabalho está em analisar o aspecto temporal das colaborações, e como estas evoluem ao longo do tempo. Este tipo de análise pode indicar não somente as novas direções da comunidade de IHC, mas também prever futuras colaborações que venham a surgir.

Outra questão seria analisar outros aspectos da rede, tais como a expansão dos grupos de pesquisa que deram origem a comunidade,

além da relação entre a diversidade de organizações envolvidas em um artigo. Também pode-se verificar a criação de escolas de pensamento, tendências nos métodos, contextos, teorias,

pesquisadores que mais disseminaram temas nos anos seguintes de suas publicações.

Centralidade de Grau		Centralidade de Proximidade		Centralidade de Intermediação	
Autor	Score	Autor	Score	Autor	Score
Cristiano Maciel	228	Raquel Oliveira Prates	0,024	Raquel Oliveira Prates	4960
Raquel Oliveira Prates	186	Cristiano Maciel	0,023	Maria C. C. Baranauskas	4163
Marcelo Soares Pimenta	180	Maria C. C. Baranauskas	0,023	Cristiano Maciel	3825
Milene Selbach Silveira	177	Clarisse Sieckenius de Souza	0,022	Junia Coutinho Anacleto	3674
Isabela Gasparini	170	Milene Selbach Silveira	0,022	Milene Selbach Silveira	3531
Carla Maria Dal S. Freitas	167	Simone Diniz J. Barbosa	0,022	Marcelo Soares Pimenta	3342
Junia Coutinho Anacleto	162	Junia Coutinho Anacleto	0,021	Simone Diniz J. Barbosa	3306
Maria C. C. Baranauskas	153	Maria Elizabeth S. Furtado	0,020	Isabela Gasparini	2087
Tayana Uchôa Conte	148	Vânia Paula de Almeida Neris	0,020	Marco Antônio A. Winckler	1996
Lucia Vilela L. Filgueiras	144	Marcelo Soares Pimenta	0,019	Maria Elizabeth S. Furtado	1960

Tabela 6. Métricas de Centralidade para a Rede de Coautoria dos Pesquisadores Prolíficos

Centralidade de Grau		Centralidade de Proximidade		Centralidade de Intermediação	
Autor	Score	Autor	Score	Autor	Score
Clarisse Sieckenius de Souza	29	Celso Alberto Saibel Santos	1	Clarisse Sieckenius de Souza	136
Maria C. C. Baranauskas	24	Heloísa Vieira da Rocha	1	Maria C. C. Baranauskas	35,5
Carla Maria Dal S. Freitas	11	Marcelo Soares Pimenta	1	Milene Selbach Silveira	18
Marcelo Soares Pimenta	10	Maria C. C. Baranauskas	0,333	Simone Diniz J. Barbosa	18
Heloísa Vieira da Rocha	8	Leonardo Cunha de Miranda	0,2	Raquel Oliveira Prates	12
Tayana Uchôa Conte	8	Clarisse Sieckenius de Souza	0,142	Jair Cavalcanti Leite	9
Simone Diniz J. Barbosa	7	Jair Cavalcanti Leite	0,833	Marcelo Soares Pimenta	4,5
Lucia Vilela L. Filgueiras	7	Milene Selbach Silveira	0,833	Heloísa Vieira da Rocha	3,5
Milene Selbach Silveira	7	Raquel Oliveira Prates	0,833	Carla Faria Leitão	3
Raquel Oliveira Prates	6	Simone Diniz J. Barbosa	0,833	Celso Alberto Saibel Santos	2

Tabela 7. Métricas de Centralidade para a Rede de Orientações de Doutorado

Centralidade de Grau		Centralidade de Intermediação	
Instituição	Score	Instituição	Score
UFRGS	90	UNICAMP	40
UNICAMP	81	UFSCar	20,5
PUC-Rio	59	UFSC	14,5
UFSCar	43	UNIFACS	10
UNIFOR	36	UEM	6,5

Tabela 8. Métricas de Centralidade para Universidades da Rede de Colaboração

Centralidade de Grau		Centralidade de Proximidade		Centralidade de Intermediação	
Autor	Score	Autor	Score	Autor	Score
Maria C. C. Baranauskas	398	Raquel Oliveira Prates	0,0243	Cristiano Maciel	4847
Cristiano Maciel	389	Maria C. C. Baranauskas	0,0238	Maria C. C. Baranauskas	4835
Marcelo Soares Pimenta	361	Cristiano Maciel	0,0232	Junia Coutinho Anacleto	4197
Carla Maria Dal S. Freitas	322	Clarisse Sieckenius de Souza	0,0227	Marcelo Soares Pimenta	3948
Junia Coutinho Anacleto	312	Junia Coutinho Anacleto	0,0227	Raquel Oliveira Prates	3597
Raquel Oliveira Prates	293	Milene Selbach Silveira	0,0227	Carla Maria Dal S. Freitas	3577
Maria Elizabeth S. Furtado	288	Simone Diniz J. Barbosa	0,0227	Maria Elizabeth S. Furtado	3466
Lucia Vilela L. Filgueiras	282	Maria Elizabeth S. Furtado	0,0222	Milene Selbach Silveira	3357
Milene Selbach Silveira	280	Marco Antônio A. Winckler	0,0212	Lucia Vilela L. Filgueiras	3249
Simone Diniz J. Barbosa	272	Vânia Paula A. Neris	0,0212	Clarisse Sieckenius de Souza	3041

Tabela 9. Métricas de Centralidade para a Rede de Colaboração Completa

AGRADECIMENTOS

Isabela Gasparini agradece ao apoio financeiro da FAPESC, Edital chamada pública FAPESC/CNPQ N° 06/2016 apoio a infraestrutura de CTI para jovens pesquisadores, projeto T.O. N°: 2017TR1755 - Ambientes Inteligentes Educacionais com Integração de Técnicas *Learning Analytics* e Gamificação.

Simone D. J. Barbosa agradece ao apoio financeiro do CNPq, processo # 309828/2015-5.

REFERÊNCIAS

- [1] Renato Balancieri. 2004. Análise de Redes de Pesquisa em uma Plataforma de Gestão em Ciência e Tecnologia: uma aplicação à Plataforma Lattes. Dissertação de Mestrado. Universidade Federal de Santa Catarina, Florianópolis, SC.

- [2] Christoph Bartneck and Jun Hu. 2009. Scientometric Analysis of the CHI Proceedings. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09), 699-708.
- [3] Mohamed Amine Chatti, Anna Lea Dyckhoff, Ulrik Schroeder, and Hendrik Thüs. 2012. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*. 4, 5/6, 318-331.
- [4] Luciano A. Digiampietri, Jesus P. Mena-Chalco, Jose J. Pérez-Alcázar, Esteban F. Tuesta, Karina V. Delgado, Rogério Mugnaini, and Gabriela S. Silva. 2012. Minerando e caracterizando dados de Currículos Lattes. In Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2012).
- [5] Luciano A. Digiampietri and Ernando E. da Silva. 2011. A framework for a social network of researchers analysis. *Iberoamerican Journal of Applied Computing*. 1, 1 (Ago 2011), 1-24.
- [6] Ademir C. Gabardo. 2015. *Análise de redes sociais: uma visão computacional*. Novatec Editora.
- [7] Isabela Gasparini, Lucas Felipe da Cunha, Marcos Hideshi Kimura, and Marcelo Soares Pimenta. 2014. Análise das redes de coautoria do simpósio brasileiro sobre fatores humanos em sistemas computacionais. In Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems (IHC '14), 323-332.
- [8] Isabela Gasparini, Felipe C. de Mendonça, Milene S. Silveira, Simone D., J. Barbosa, and Rebeca Schroeder. 2017. Crossing the borders of IHC: where else have our researchers been publishing? In Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems (IHC 2017). ACM, New York, NY, USA, Article 19, 10 pages. DOI: <https://doi.org/10.1145/3160504.3160547>
- [9] Jennifer Golbeck. 2015. *Introduction to Social Media Investigation: A Hands-On Approach*. Elsevier.
- [10] Aziz Galvão da Silva Júnior. 2016. Análise de Redes Sociais aplicada à Assistência Técnica e Extensão Rural (ATER). Acessado em 30 de Outubro de 2016 de <http://pt.slideshare.net/equipeagropius/agropius-social-networkanalysis>
- [11] William Takahiro Maruyama and Luciano Antonio Digiampietri. 2016. Co-authorship prediction in academic social network. In Proceedings of the V Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2016), 79-90.
- [12] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank Citation Ranking: bringing order to the web*. Stanford InfoLab: Technical Report.
- [13] Jesús Pascual Mena-Chalco and Roberto Marcondes Cesar Junior. 2009. Scriptlattes: an open-source knowledge extraction system from the lattes platform. *Journal of the Brazilian Computer Society*. 15, 4, 31-39.
- [14] Mark E. J. Newman. 2004. *Who is the best connected scientist a study of scientific coauthorship networks*. Springer Berlin Heidelberg.
- [15] Fábio Ricotta. 2016. O que é PageRank? Acessado em 30 de Outubro de 2016 de <http://www.agenciamestre.com/marketing-digital/o-que-e-pagerank/>
- [16] John Scott. 2000. *Social network analysis: a handbook*. Sage Publications.
- [17] Alzira Karla Araújo da Silva, Ricardo Rodrigues Barbosa, and Emeide Nóbrega Duarte. 2012. Rede social de coautoria em ciência da informação: estudo sobre a área temática de "organização e representação do conhecimento". *Informação & Sociedade*. 22, 2 (maio/ago. 2012), 63-79.
- [18] Igor Steinmacher, Marco Aurélio Graciotto Silva, Filipe Roseiro Cogo, Ana Paula Chaves, and Marco Aurélio Gerosa. 2013. Tópicos de Pesquisa e Rede de Coautoria no Simpósio Brasileiro de Sistemas Colaborativos. In: Proceedings of X Brazilian Symposium on Collaborative Systems (SBSC'13), 88.
- [19] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2005. *Introduction to Data Mining*, Addison-Wesley.
- [20] Shahadat Uddin, Liaquat Hossain, Alireza Abbasi, Kim Rasmussen. 2012 Trend and efficiency analysis of coauthorship network. *Scientometrics*. 90, 2 (February 2012), 687-699.
- [21] Joel Weisz and M.C. Roco. 1996. *Redes de pesquisa e educação em engenharia nas américas*. FINEP.
- [22] Evelien Otte and Ronald Rousseau. 2002. Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, 28(6), 441-453.