



**Universidade de Brasília**

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# Análise de Redes Sociais Científicas para Ampliação de Redes de Colaboração

Aurélio Ribeiro Costa

Dissertação apresentada como requisito parcial  
para conclusão do Mestrado em Informática

Orientadora  
Prof.<sup>a</sup> Dr.<sup>a</sup> Célia Ghedini Ralha

Brasília  
2016

Universidade de Brasília — UnB  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Mestrado em Informática

Coordenadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Célia Ghedini Ralha

Banca examinadora composta por:

Prof.<sup>a</sup> Dr.<sup>a</sup> Célia Ghedini Ralha (Orientadora) — CIC/UnB  
Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Emilia Machado Telles Walter — CIC/UnB  
Prof. Dr. Victor Ströele de Andrade Menezes — DCC/UFJF

### **CIP — Catalogação Internacional na Publicação**

Costa, Aurélio Ribeiro.

Análise de Redes Sociais Científicas para Ampliação de Redes de Colaboração / Aurélio Ribeiro Costa. Brasília : UnB, 2016.

123 p. : il. ; 29,5 cm.

Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2016.

1. Análise de Redes Sociais, 2. Design Science Research, 3. NoSQL,  
4. Sistemas de Recomendação

CDU 004

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil

**Aurélio Ribeiro Costa**

**Análise de Redes Sociais Científicas para Ampliação de Redes de Colaboração**

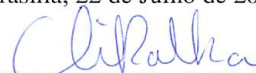
Dissertação aprovada como requisito parcial para obtenção do grau de **Mestre** no Curso de Pós-graduação em Informática da Universidade de Brasília, pela Comissão formada pelos professores:

Orientador: Prof.<sup>a</sup> Dr.<sup>a</sup> Célia Ghedini Ralha, CIC/UNB

Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Emília Machado Telles Walter, CIC/UNB

Prof. Dr. Victor Ströele de Andrade Menezes, UFJF

Vista e permitida a impressão.  
Brasília, 22 de Julho de 2016.



Prof.<sup>a</sup> Dr.<sup>a</sup> Célia Ghedini Ralha  
Coordenadora do Programa de Pós-Graduação em Informática  
Departamento de Ciência da Computação  
Universidade de Brasília.

Universidade de Brasília

# Dedicatória

Dedico a cada gota de suor que se perdeu nessa caminhada.

# Agradecimentos

Agradeço primeiro a *Deus* por sua mão invisível que, mesmo que não compreendamos a tempo, coloca tudo no seu devido lugar.

Agradeço aos meus pais *Sônia* e *Álvaro* por além de terem me colocado no mundo, terem garantido o alicerce que me fez chegar até aqui.

Agradeço minha esposa *Andressa* pelo amor e suporte incondicional em todos os passos dessa caminhada.

Agradeço *Prof.<sup>a</sup> Dr.<sup>a</sup> Célia Ghedini Ralha* por ter acreditado que esse trabalho seria viável e por sua presença sempre constante ao longo desse árduo caminho.

Agradeço a *Dr.<sup>a</sup> Vanessa* por suas sempre pertinentes opiniões e sugestões durante as reuniões de orientação e nos encontros do InfoKnow.

Agradeço aos meus colegas de trabalho no Supremo Tribunal Federal pelo apoio dado durante minha licença para conclusão desse trabalho.

# Resumo

Redes de colaboração científicas são redes sociais com foco na condução de pesquisas científicas. A ampliação dessas redes é um dos fatores que pode ampliar as parcerias podendo resultar na melhoria das pesquisas desenvolvidas por grupos de pesquisa no Brasil e no mundo. Dentre as abordagens existentes para ampliação dessas redes de colaboração estão os sistemas de recomendação. Nesse trabalho, foi explorada a análise das redes de colaboração existentes entre pesquisadores vinculados a programas de pós-graduação de universidades federais no país, para sugerir novas parcerias, através da classificação das áreas dos periódicos publicados e do cálculo dos centróides (dos vértices com mais ligações) dessas áreas. A metodologia utilizada neste trabalho centra-se no *Design Science Research* para guiar a elaboração do modelo e a construção do artefato de recomendação para fins de validação. Para realização do estudo de caso foram coletados dados de publicações no repositório digital da área de Ciência da Computação denominado *Digital Bibliography & Library Project* (DBLP), os quais foram modelados em um banco de dados NoSQL orientado a grafo. Os dados coletados são referentes às publicações dos pesquisadores vinculados aos Programas de Pós-Graduação na área de Ciência da Computação da UFMG, USP, UFRN e UnB, sendo considerados também pesquisadores não vinculados a esses programas mas que possuem relacionamentos de coautoria. No contexto de redes de colaboração científica, a análise de rede pode ser utilizada para avaliar os relacionamentos já formados, bem como para estimular a formação de novas parcerias. Os resultados das recomendações foram avaliados através de um questionário online. Neste questionário ficou evidenciado que as recomendações produzidas pelo artefato desenvolvido alcançaram um percentual de aceitação de 31%, considerado bom, uma vez que na literatura da área os sistemas de recomendação alcançam uma média abaixo de 30%.

**Palavras-chave:** Análise de Redes Sociais, Design Science Research, NoSQL, Sistemas de Recomendação

# Abstract

Scientific collaboration networks are social networks focused on scientific research. The broadening of these networks is a factor that may improve the researches developed by research groups in Brazil and in the world. Among the existent approaches to broadening these collaboration networks one of them is partnership recommendation. This research explores the analysis of the existent collaboration networks of researchers, members of graduate programs of Federal universities in Brazil, for suggesting new partnership through the classification of research areas of published papers and the centroid (most linked vertices) computation for those areas. This work uses the Design Science Research methodology to develop the analysis model and to build a recommendation artifact for validating the model. The study case conducted to validate the analysis model and recommendation artifact uses a digital repository of publications named Digital Bibliography & Library Project (DBLP), its data was modeled in a NoSQL Graph Database starting from researchers members of graduated programs in Computer Science and Informatics of UFMG, USP, UFRN and UnB, considering researchers non-members to those programs but that make part of coauthoring relationships. Considering the scientific collaboration networks context, network analysis may be used to evaluate current relationships as well as stimulate the creating of new relationships. The results of recommendations were evaluated by an online questionnaire which showed the recommendations generated by the developed artifact reached an acceptance rate of 31%, what was considered reasonable for a recommendation system, besides the recommendation systems literature show a mean lower than 30%.

**Keywords:** Social Network Analysis, Design Science Research, NoSQL, Recommendation Systems

# Sumário

<b>Lista de Figuras</b>	<b>xi</b>
<b>Lista de Tabelas</b>	<b>xii</b>
<b>1 Introdução</b>	<b>2</b>
1.1 Motivação e Problema . . . . .	3
1.2 Objetivos . . . . .	4
1.3 Hipótese . . . . .	4
1.4 Metodologia . . . . .	4
1.5 Apresentação do Documento . . . . .	5
<b>2 Fundamentos</b>	<b>6</b>
2.1 <i>Design Science Research</i> . . . . .	6
2.2 Recuperação da Informação . . . . .	12
2.3 Redes de Colaboração . . . . .	14
2.4 Sistemas de Recomendação . . . . .	18
<b>3 Trabalhos Correlatos</b>	<b>22</b>
3.1 Revisão <i>Quasi</i> -sistemática . . . . .	22
3.2 Principais Artigos . . . . .	25
<b>4 Solução Proposta</b>	<b>28</b>
4.1 Modelo Conceitual . . . . .	28
4.2 Modelagem de dados . . . . .	32
4.3 Detalhamento da Implementação . . . . .	33
4.4 Tecnologias Utilizadas . . . . .	36
<b>5 Experimentação e Resultados</b>	<b>42</b>
5.1 Estudo de Caso . . . . .	42
5.2 Análise dos Resultados . . . . .	43
<b>6 Conclusões e Trabalhos Futuros</b>	<b>47</b>
<b>Referências</b>	<b>49</b>
<b>A Questionário de Verificação de Recomendações</b>	<b>55</b>
<b>B Trecho da Base de Dados DBLP</b>	<b>58</b>



C	Áreas do Sistema de Classificação da ACM	61
D	Scripts de Coleta dos Pesquisadores	64
E	Script Bag of Areas	68
F	Script de Geração das Recomendações	72
G	Questionários de Avaliação	77

# Lista de Figuras

2.1	<i>Framework</i> conceitual para pesquisas em SI usando <i>Design Science Research</i> (tradução literal de Simon (1996)). . . . .	7
2.2	Precisão e Revocação em um conjunto de documentos . . . . .	13
2.3	Grafo direcionado com dois vértices A e B e uma aresta direcionada de A para B, com peso $w = 300$ . . . . .	15
2.4	Distribuição da quantidade de publicações por autor . . . . .	17
2.5	Rede de colaboração centrada um pesquisador . . . . .	17
3.1	Repositórios de publicações usados. . . . .	24
3.2	Resultado da fase de extração das publicações. . . . .	25
3.3	Frequência dos critérios usados pelas publicações aceitas. . . . .	26
3.4	Frequência dos critérios usados pelas publicações rejeitadas . . . . .	26
4.1	<i>Workflow</i> da Solução Proposta. . . . .	29
4.2	Distribuição dos tipos de publicações presentes na DBLP . . . . .	30
4.3	Trecho do Sistema de Classificação da ACM para áreas da Computação. . . . .	31
4.4	Modelo de dados usado para representar as entidades do modelo de recomendação . . . . .	33
4.5	Processo de carga e visualização . . . . .	34
4.6	O processo de pré-processamento . . . . .	35
4.7	Processo de geração e validação das recomendações . . . . .	35
5.1	Evolução anual das métricas de avaliação das recomendações . . . . .	44
5.2	Nuvem de palavras das áreas de pesquisa indicadas . . . . .	46

# Lista de Tabelas

2.1	Métodos de avaliação do <i>Design Science Research</i> . . . . .	9
4.1	Estados de periódicos conforme documento de área da CAPES (2013) . . . . .	31
4.2	Resultado da consulta de coautores e a quantidade de publicações . . . . .	37
4.3	Cálculos estatísticos gerados pelo Gephi . . . . .	41
5.1	Exemplo de recomendações de parceria . . . . .	43
5.2	Resultado das métricas de avaliação das recomendações . . . . .	44
5.3	Síntese das respostas ao questionário de avaliação . . . . .	45

# Lista de Códigos

4.1	Retorna os coautores e a quantidade de publicações . . . . .	37
4.2	Cálculo de fluxo máximo utilizando NetworkX . . . . .	38
4.3	Criação da Árvore Geradora Mínima de um Grafo G pela NetworkX . . . . .	38
4.4	Exemplo de filtro de pesquisadores . . . . .	39
D.1	Script para coleta dos dados dos pesquisadores da UFMG . . . . .	64
D.2	Script para coleta dos dados dos pesquisadores da USP . . . . .	64
D.3	Script para coleta dos dados dos pesquisadores da UFRN . . . . .	65
D.4	Script para coleta dos dados dos pesquisadores da UnB . . . . .	66
E.1	Script para criação do grupo de áreas de pesquisa . . . . .	68
F.1	Script para geração das recomendações . . . . .	72

# Abreviaturas

**ACM** - *Association for Computing Machinery*

**DBLP** - *Digital Bibliography & Library Project*

**CAPES** - *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*

**CNPq** - *Conselho Nacional de Desenvolvimento Científico e Tecnológico*

**DOM** - *Document Object Model*

**IEEE** - *Institute of Electrical and Electronics Engineers*

**JSON** - *Javascript Object Notation*

**NLTK** - *Natural Language ToolKit*

**NoSQL** - *Not only Structured Query Language*

**SAX** - *Simple API for XML*

**RI** - *Recuperação da Informação*

**SI** - *Sistemas de Informação*

**SR** - *Sistemas de Recomendação*

**TI** - *Tecnologia da Informação*

**UFMG** - *Universidade Federal de Minas Gerais*

**UFRN** - *Universidade Federal do Rio Grande do Norte*

**UnB** - *Universidade de Brasília*

**USP** - *Universidade de São Paulo*

**Web** - *World Wide Web*

**XML** - *Extensible Markup Language*

# Capítulo 1

## Introdução

No âmbito da ciência, a imagem do pesquisador isolado faz parte do passado, pois atualmente o processo de produção científica requer associações, negociações e estratégias para interligar o maior número possível de pesquisadores. O compartilhamento de informações, a junção de competências e a união de esforços de pesquisadores na busca de metas em comum impulsionam a produção de conhecimento (Balancieri et al., 2005). O trabalho compartilhado proporciona economia de tempo e de recursos financeiros e materiais, e, portanto, é também estimulado pelas agências financiadoras de pesquisa. Todos esses fatores contribuem para que atualmente seja bastante valorizado o pesquisador capaz de formar boas, eficientes e produtivas equipes de trabalho. Além disso, o avanço das tecnologias de comunicação e as facilidades de deslocamento vêm proporcionando um aumento no número de estudos realizados de modo compartilhado e por meio de redes de colaboração, tanto entre indivíduos como entre instituições e países (Maia et al., 2008).

Independente do status dos indivíduos participantes de um grupo de pesquisa, a transferência do conhecimento e das habilidades destaca-se como um dos maiores benefícios da colaboração (Katz and Martin, 1997). Em equipe, o *background* individual de cada pesquisador é socializado no grupo, facilitando a transferência de conhecimentos e habilidades, além de possibilitar o aprendizado de todos os participantes e estimular a criatividade e as novas ideias. Discutir diferentes pontos de vista pode gerar novas perspectivas, característica que se amplia quando os colaboradores são oriundos de diferentes áreas (Vanz and Stumpf, 2010). Nesse contexto, as recomendações de parcerias em redes de colaboração pode ser um instrumento valioso no fortalecimento das pesquisas realizadas por um grupo de pesquisa.

Conforme Barabási et al. (2002), para um novo autor, que aparece pela primeira vez em uma publicação, é mais provável que essa publicação seja realizada em coautoria com algum autor que já possua um grande número de coautores (ligações) do que com alguém menos conectado. Como resultado, autores “antigos” com mais ligações aumentarão suas quantidades de coautorias a uma taxa maior que aqueles com poucas ligações. A esse fenômeno Barabási and Albert (1999) denominaram anexação preferencial. Os modelos de recomendação de parcerias em redes de colaboração que se baseiam na topologia da rede, como o aqui apresentado, seguem esse fenômeno ao indicar pesquisadores com maior número de publicações em suas áreas de pesquisa.

## 1.1 Motivação e Problema

Na condução de uma pesquisa por um grupo de pesquisadores é fácil determinar quem são os especialistas na mesma área de atuação. No entanto, essa facilidade deixa de existir, ou é reduzida, quando é necessário interagir com outros grupos ou até com outras áreas de pesquisa. Essa dificuldade em se determinar os especialistas em uma área pode gerar perda de rendimento na pesquisa causando atraso na obtenção de resultados ou até a não conclusão de estudos.

Ao mesmo tempo que cresce o número de estudos e publicações científicas compartilhadas pelo mundo, aumenta o interesse em analisar essas colaborações, que podem apresentar diferentes características e variadas motivações. Independentemente das especificidades, é possível afirmar que a coautoria de artefatos gerados pela atividade científica, particularmente de publicações, é um indicador de colaboração (Katz and Martin, 1997; Newman, 2004; Cronin, 2005; Kim, 2006). Estudos sobre coautoria têm sido empregados, por exemplo, para verificar diferenças entre colaborações acadêmicas e técnicas (Meyer and Bhattacharya, 2004), características de colaboração em determinadas disciplinas (Mählck and Persson, 2000; Garg and Padhi, 2001; Glanzel, 2002; Yoshikane and Kageura, 2004; Kim, 2006), colaboração entre pesquisadores de mesmas ou diferentes instituições ou países (Mählck and Persson, 2000; Glänzel, 2001; Tomizawa and Shirabe, 2002; Alonso-Arroyo et al., 2005; Kim, 2006). Os resultados desses estudos revelam que as colaborações entre autores têm aumentado em todas as áreas de ciência. Entretanto, o grau de colaboração difere entre elas, ou seja, nas ciências naturais, por exemplo, não só o número de artigos em coautoria como o número de autores por artigo é maior que nas ciências sociais (Katz and Martin, 1997; Meadows and de Lemos, 1999; Yoshikane and Kageura, 2004). Os estudos sobre coautoria e colaboração mostram que geralmente trabalhos teóricos geram artigos com menos autores do que aqueles de natureza experimental (Katz and Martin, 1997; Cronin, 2005). Além disso, demonstram que os trabalhos cooperativos internacionais têm maior impacto e visibilidade (Glanzel, 2002), e que a colaboração aumenta a produtividade dos pesquisadores (Maia et al., 2008).

Diante do cenário exposto, a motivação para a condução dessa pesquisa é fundamentada na relevância da solução do problema de se determinar quem são os especialistas de uma determinada área de pesquisa. Motiva ainda essa pesquisa a geração de recomendações de parcerias em pesquisas científicas, visando a ampliação das redes de colaboração dos pesquisadores, utilizando como método a análise da rede de colaboração formada pelos relacionamentos de coautoria em publicações realizadas.

Tanto quanto sabemos, não há ferramenta com dados de pesquisadores no Brasil que produzam recomendações de parcerias em pesquisas científicas. Para mitigar a dificuldade de encontrar especialistas em uma área de atuação, uma alternativa é o desenvolvimento de um artefato que recomende pesquisadores ou outros grupos de pesquisa que se destaquem nas áreas de interesse a partir da rede social formada pelas publicações anteriores de cada pesquisador. Tal recomendação pode ser realizada na forma de sugestões de novos pesquisadores para compor um grupo de pesquisa ou de um grupo de pesquisa com o qual se possa formar parcerias. O objetivo de tal artefato é auxiliar no direcionamento de ações objetivando a melhoria dos indicadores de produção científica.

A questão de pesquisa que norteia este trabalho é a possibilidade de se definir e

implementar um artefato para análise de redes de colaboração científica que realize recomendações de parcerias meritorias para pesquisadores.

## 1.2 Objetivos

Esta pesquisa tem como objetivo geral desenvolver um modelo de recomendação de parceria para pesquisas científicas e implementá-lo em um artefato utilizando informações de grupos de pesquisadores na área de Ciência da Computação, vinculados a determinados programas de pós-graduação em universidades federais do país.

Como objetivos secundários cita-se:

- construir de um banco de dados NoSQL baseado em grafo que armazene a rede de colaboração científica de pesquisadores;
- gerar cálculos analíticos da rede de colaboração armazenada para análise de autores e coautores;
- avaliar o artefato gerado através de um questionário *online* para ajustes e melhorias no artefato;
- publicar em veículos da área de Ciência da Computação os resultados dessa pesquisa para divulgação à comunidade do artefato desenvolvido.

## 1.3 Hipótese

A disposição em colaborar e o compartilhamento eficiente de informação entre os atores de uma rede social científica asseguram ganhos, porque cada participante melhora, valendo-se das informações às quais passa a ter acesso e que poderão reduzir as incertezas e promover o crescimento mútuo. Segundo Yu et al. (2001), cada ator tem muita informação sobre sua situação, mas não tem informação sobre outras situações. Para reduzir a incerteza e consolidar a parceria, os atores precisam ter mais informações confiáveis de seus parceiros. Assim todos ganham, porque cada ator vai construir alicerces e desenvolver novas ações tendo como base as informações compartilhadas (Tomaél and Marteleto, 2007).

Com a finalidade de promover o compartilhamento de informação e consolidar parcerias através da descoberta de especialistas, a hipótese considerada nesta pesquisa é a possibilidade de serem geradas recomendações de possíveis parcerias válidas para os pesquisadores participantes de uma rede social científica. Espera-se que a aceitação das recomendações seja superior a apresentada na literatura acerca da avaliação de sistemas de recomendação em redes de colaboração, que apresenta uma média de aceitação de 29%, conforme descrito por Debnath et al. (2008).

## 1.4 Metodologia

A abordagem utilizada para construção desse modelo foi baseada na análise da rede de colaboração científica formada pelos pesquisadores da área de Computação, rede essa composta pelas relações de coautoria em publicações. Como estudo de caso para avaliar



o modelo de recomendação desenvolvido, foram coletados dados de publicações em periódicos de pesquisadores vinculados a programas de pós-graduação na área de Computação de quatro universidades brasileiras: Universidade Federal de Minas Gerais (UFMG), Universidade de São Paulo (USP), Universidade Federal do Rio Grande do Norte (UFRN) e Universidade de Brasília (UnB). Esses programas foram escolhidos segundo avaliação na CAPES com conceitos de sete a quatro, respectivamente. A fonte de publicações utilizada foi o banco de dados de publicações da área de Ciência da Computação denominado *Digital Bibliography & Library Project* (DBLP), cujo número de publicações cadastradas é de 3.369.107 com 1.731.100 autores. Partindo do conjunto inicial de 139 autores de publicações vinculados aos programas de pós-graduação em Computação citados, foram filtrados 3.158 autores de 3.056 publicações. Vale salientar que sempre que uma publicação era tratada, todos os coautores eram incluídos no banco dados NoSQL da rede de colaboração.

Uma limitação encontrada no decorrer da pesquisa foi a identificação automática das áreas de pesquisa dos pesquisadores por meio dos nomes dos periódicos utilizados com veículos para as publicações científicas. Essa limitação é um problema, uma vez que a determinação das áreas de pesquisa do pesquisador é um dos pilares para a realização das recomendações de parceria. Como exemplo podemos citar o periódico da ACM *Journal of Experimental Algorithmics* cujas áreas de interesse declaradas pelo periódico incluem *Design and analysis of algorithms* e *Data structures*. Por outro lado temos o *Journal of Universal Computer Science* que na sua descrição de escopo afirma que lida com todos os aspectos da Ciência da Computação. Esses exemplos ilustram o problema de classificação automática das áreas de pesquisa baseados na denominação dos periódicos, problema que deve ser melhor trabalhado em pesquisas futuras.

## 1.5 Apresentação do Documento

A presente dissertação está organizada da seguinte forma: no Capítulo 2, são apresentados os fundamentos teóricos que nortearam a pesquisa; no Capítulo 3, é descrito o estado da arte dos conceitos explorados nesta pesquisa, através dos trabalhos correlatos; no Capítulo 4, o modelo desenvolvido é detalhado; no Capítulo 5, é apresentada a implementação do modelo na forma de um artefato de software e os resultados obtidos através da condução do estudo de caso com a avaliação do modelo e do artefato através da aplicação de um questionário *online* para uma amostra da população total de pesquisadores utilizada na pesquisa; e no Capítulo 6, são apresentadas as conclusões e os trabalhos futuros derivados dessa pesquisa.

# Capítulo 2

## Fundamentos

Essa pesquisa é fundamentada na análise das Redes de Colaboração Científica, ou seja, na análise das interações sociais que são realizadas por agentes na publicação de suas pesquisas. Outro aspecto da pesquisa envolve os sistemas de recomendação, importantes ferramentas no auxílio às tomadas de decisão, aqui usadas para auxílio na seleção de parcerias em pesquisas científicas. Uma teoria importante para auxiliar nos fundamentos citados está sustentada nos sistemas de recuperação da informação empregados na coleta e avaliação da qualidade dos dados recuperados. Como metodologia geral adotada na condução dessa pesquisa, o *Design Science Research*, que defende a construção de um artefato para validar a teoria que está sendo proposta, sendo muito comum como modelo conceitual de pesquisa na área de Sistemas de Informação (SI).

Esse capítulo apresenta os aspectos tóricos adotados nessa pesquisa, descrevendo cada fundamento e a forma como foi adotado. Na Seção 2.1, é apresentada a metodologia adotada para a condução da pesquisa; na Seção 2.2, noções de recuperação da informação são introduzidas e são apresentadas as métricas para avaliação da recuperação da informação; na Seção 2.3 os conceitos de redes de colaboração são apresentados; e na Seção 2.4 os sistemas de recomendação são descritos, assim como o conceito e os métodos de similaridade são apresentados.

### 2.1 *Design Science Research*

*Design* é essencialmente um processo de busca por uma solução efetiva para um problema. A solução de um problema pode ser vista como o uso dos meios disponíveis para atingir os fins desejados, enquanto satisfaz as leis existentes no ambiente, conforme descrito por Simon (1996). A abstração e representação de meios, fins e leis apropriadas são componentes cruciais do *Design Science Research*. A Figura 2.1 apresenta o *framework* conceitual para pesquisas que adotem o *Design Science Research* para SI.

*Design Science Research* com frequência simplifica um problema representando explicitamente apenas um subconjunto de meios, fins e leis relevantes ou decompondo um problema em subproblemas menores. Tais simplificações e decomposições podem não ser realistas o suficiente para ter um impacto significativo na prática, porém podem representar um ponto inicial. O progresso da pesquisa é realizado iterativamente com a expansão do seu escopo. Com o refinamento de meios, fins e leis mais realistas, o desenvolvimento do artefato se torna mais relevante e valioso.

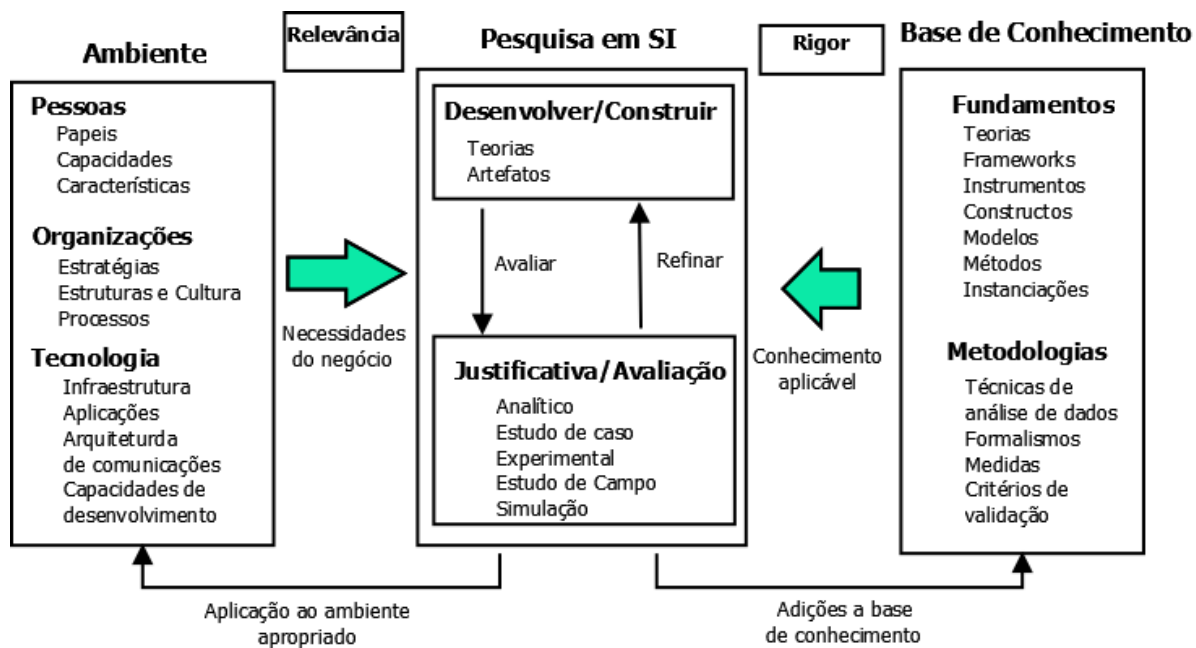


Figura 2.1: *Framework* conceitual para pesquisas em SI usando *Design Science Research* (tradução literal de Simon (1996)).

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. A aplicação de *Design Science Research* nesta pesquisa tem o objetivo de construir um artefato que modele as parcerias entre pesquisadores através de relações de coautoria em publicações científicas.

Uma conjectura teórica adotada neste trabalho e apresentada na Seção 1.3, é a de que pesquisadores com uma rede de colaboração congregando aspectos de tamanho e qualidade estejam aptos a ter uma produção científica qualificada. Buscando validar ou refutar essa conjectura, foi construído um grafo de relacionamento baseado na coautoria entre pesquisadores.

Segundo proposto por Hevner et al. (2004), existem sete regras que uma pesquisa deve seguir para implementar o *Design Science*. Descrevemos as regras e de que forma esta pesquisa as implementa.

1. **Design como um artefato:** O resultado do *Design Science Research* em SI é, por definição, um artefato de Tecnologia da Informação (TI) que tenha o propósito de resolver um importante problema organizacional. Ele deve ser descrito efetivamente, permitindo sua implementação e aplicação a um domínio apropriado.

Orlikowski and Iacono (2001) consideram o artefato de TI o *core subject matter* do campo de SI. Embora eles articulem diversas definições do termo Artefato de TI, muitas das quais incluem componentes da organização e pessoas envolvidas no uso de um artefato baseado em computador, eles enfatizam a importância desses conjuntos de propriedades culturais empacotados em alguma forma socialmente reconhecida

como hardware e software, ou seja, o Artefato de TI como uma instanciação. Weber (1987) argumenta que teorias baseadas em artefatos de longa vida são fundamentais para a disciplina de SI, definição essa reiterada por Weber (2003). Essas teorias devem explicar como artefatos são criados e adaptados segundo as mudanças dos ambientes nos quais os artefatos estão inseridos, assim como as tecnologias que suportam esses artefatos.

O artefato desenvolvido nesta pesquisa consiste em uma ferramenta automatizada. Tal artefato utiliza um modelo de recomendação baseado em análise de redes sociais científicas para gerar sugestões de novas parcerias em pesquisa científicas.

2. **Relevância do problema:** O objetivo da pesquisa em computação é a aquisição de conhecimento e entendimento que permitam o desenvolvimento e a implementação de soluções baseadas em tecnologia para lidar com importantes problemas de negócio não resolvidos. As ciências comportamentais abordam esse objetivo através do desenvolvimento e justificação de teorias que expliquem ou prevejam a ocorrência de fenômenos. *Design Science* aborda esse objetivo através da construção de artefatos inovadores objetivando a mudança nos fenômenos. Cada um deve informar e desafiar o outro.

Formalmente, um problema pode ser definido como a diferença entre o objetivo e o estado atual de um sistema. A solução de problema pode ser definida como um processo de busca usando ações para reduzir ou eliminar as diferenças (Simon, 1996).

A relevância de qualquer esforço para o *Design Science Research* é considerada no escopo de uma comunidade constituída. Para pesquisadores em SI, a comunidade é constituída por praticantes que planejam, gerenciam, projetam, implementam, operam e avaliam as tecnologias que permitem seu desenvolvimento e implementação. Para ser relevante para essa comunidade, a pesquisa deve ter foco nos problemas enfrentados e nas oportunidades enxergadas pela interação de pessoas, organizações e TI. Essa comunidade agradecerá artefatos que permitem que os problemas sejam enfrentados.

A relevância desta pesquisa consiste em desenvolver um artefato que facilite a identificação de especialistas em uma área de pesquisa diferente da que se costuma atuar e, além disso, que recomende pesquisadores que se destaquem nessas áreas.

3. **Avaliação da pesquisa:** A utilidade, qualidade, eficácia de um artefato deve ser rigorosamente demonstrada via métodos de avaliação bem executados. Avaliação é um componente primordial no processo de pesquisa. O ambiente de negócio estabelece os requisitos nos quais a avaliação do artefato será baseada. Este ambiente inclui a infraestrutura técnica a qual por si só é incrementalmente construída por meio da implementação de novos artefatos. Dessa forma, a avaliação inclui a integração do artefato dentro da infraestrutura do ambiente de negócio. Os métodos de avaliação descritos são os informados na Tabela 2.1.

A avaliação da pesquisa foi realizada durante a construção do artefato através dos métodos de análise estática, com o objetivo de verificar a complexidade do artefato, e de análise dinâmica, pelo cálculo das métricas de recuperação da informação precisão (*precision*), revocação (*recall*) e média harmônica (*f<sub>1</sub> measure*). Foi utilizado

também o método de observação por estudo de caso através da disponibilização de um questionário online a uma amostra da população de pesquisadores participantes da rede de colaboração armazenada no banco de dados NoSQL.

Tabela 2.1: Métodos de avaliação do *Design Science Research*.

Método de Avaliação	Descrição
Observação	Estudo de Caso: Estuda o artefato em profundidade no ambiente de aplicação
	Estudo de Campo: Monitora o uso do artefato em múltiplos projetos
Analítico	Análise Estática: Examina a estrutura do artefato para qualidades estáticas (Ex: complexidade)
	Análise Arquitetural: Estuda a aderência técnica do artefato à arquitetura técnica do SI
	Otimização: Demonstrar propriedades ótimas inerentes ao artefato ou prover limites de otimalidade no comportamento do artefato
	Análise dinâmica: Estuda o artefato em uso para qualidades dinâmicas (Ex: desempenho)
Experimental	Experimento Controlado: Estuda o artefato em um ambiente controlado por qualidades (Ex: usabilidade)
	Simulação: Executa o artefato com informações artificiais
Teste	Funcional (Caixa Preta): Executa as interfaces do artefato para descobrir falhas e identificar defeitos
	Estrutural (Caixa Branca): Executa testes de cobertura de alguma métrica na implementação do artefato
Descritivo	Argumento Informado: Usa informação de base de conhecimento para construir um argumento convincente da utilidade do artefato
	Cenários: Constrói cenários detalhados sobre o artefato para demonstrar sua utilidade

4. **Contribuições da Pesquisa:** O *Design Science Research* efetivo deve prover claras contribuições para as áreas exploradas pelo artefato, conhecimento do projeto de construção (ex: fundamentos) e/ou conhecimento sobre a avaliação do projeto (ex: metodologias). *Design Science Research* possui o potencial para três tipos de contribuições de pesquisa baseadas em inovação, generalização e significância do artefato desenvolvido. Uma ou mais das seguintes contribuições devem se encontradas em um projeto de pesquisa.

- (a) Artefato projetado: Mais frequentemente a contribuição do *Design Science Research* é um artefato por si só. O artefato deve permitir a solução de problemas relevantes. Ele pode estender a base de conhecimento ou aplicar um conhecimento existente de maneiras inovadoras. Metodologias de desenvolvimento de sistemas, ferramentas de projetos, protótipos de sistemas são exemplos de tais artefatos.
- (b) Fundamentos: O desenvolvimento criativo de construtos, modelos, métodos ou instâncias, apropriadamente avaliados, que estendam e melhorem os fundamentos existentes na base de conhecimento do *Design Science Research* também são importantes contribuições.

- (c) Metodologias: O desenvolvimento criativo de métodos de avaliação e novas métricas de avaliação proveem contribuições do *Design Science Research*. Métricas de medições e avaliações em particular são componentes cruciais do *Design Science Research*.

Pode-se citar como contribuições produzidas por esta pesquisa: (i) um modelo de recomendação baseado em redes de colaboração científica, (ii) um artefato que implementa o modelo de recomendação, e (iii) um banco de dados NoSQL em grafo com os dados de pesquisadores de programas de pós-graduação em Ciência da Computação das universidades federais UFMG, USP, UFRN e UnB.

5. **Rigor da avaliação:** Rigor diz respeito a forma pela qual a pesquisa é conduzida. O *Design Science Research* requer a aplicação de métodos rigorosos em ambos, construção e avaliação do artefato projetado. Na pesquisa usando a ciência comportamental, o rigor é frequentemente medido pela aderência a apropriada coleta de informação e técnica de análise. Com frequência, *Design Science Research* usa o formalismo matemático para descrever o artefato especificado e desenvolvido. Entretanto, os ambientes nos quais os artefatos de TI devem executar e os artefatos em si podem dispensar um excessivo formalismo. Ou, em uma tentativa de ser matematicamente rigoroso, importantes partes do problema podem ser abstraídas ou tomadas por irrelevantes. Em particular, especificamente referente a atividade de construção, o rigor deve ser avaliado levando-se em consideração a aplicabilidade e a generalização do artefato.

Foi definido um índice de aceite das recomendações feitas aos pesquisadores, o qual foi ajustado conforme avaliação das recomendações fornecidas;

6. **Design como um processo de busca:** O *Design Science Research* é inerentemente iterativo. A busca pelo melhor projeto ou pelo ótimo é, com frequência, intratável para problemas reais em computação. Estratégias de busca que usam heurísticas produzem resultados viáveis, bons e que podem ser implementados no ambiente da solução. *Design* é essencialmente um processo de busca de uma solução efetiva para um problema. A solução de problemas pode ser vista como o uso dos meios disponíveis para se alcançar os fins desejados, ao mesmo tempo que satisfazem leis existentes no ambiente (Simon, 1996). A abstração e representação dos meios, fins e leis apropriados são componentes cruciais do *Design Science Research*. Esses fatores são dependentes do problema e do ambiente e invariavelmente envolvem criatividade e inovação. Meios são o conjunto de ações e recursos disponíveis para se construir uma solução. Fins representam objetivos e restrições da solução. Leis são forças incontrolláveis do ambiente. *Design* efetivo requer conhecimento tanto do domínio da aplicação quanto do domínio da solução.

*Design Science Research* frequentemente simplifica um problema explicitamente, representando apenas um subconjunto dos meios, fins e leis relevantes ou pela decomposição de um problema em subproblemas mais simples. Tais simplificações e decomposições podem não ser realistas o suficiente para ter um impacto significativo na prática, mas podem representar um ponto de partida. O progresso é feito iterativamente através da expansão do escopo do problema de projeto. Assim, meios, fins e leis são refinadas e tornadas mais realistas e o artefato de projeto se

torna mais relevante e valoroso. Meios, fins e leis para um problema em SI podem ser frequentemente representados usando ferramentas de pesquisa matemáticas operacionais. Meios podem ser representados por variáveis de decisão cujos valores constituem uma solução implementável. Fins são representados usando uma função de utilidade e restrições que podem ser expressas em termos de variáveis de decisão e constantes. Leis são representadas pelo valor de constantes usadas na função de utilidade e restrições.

O conjunto de possíveis soluções para qualquer problema é especificado como todos os possíveis meios que satisfaçam todas as possíveis condições finais consistentes com as leis identificadas. Quando isso puder ser formulado apropriadamente e expressado matematicamente, técnicas de operação de pesquisa padrão podem ser usadas para determinar uma solução ótima para as condições finais especificadas. Dada a natureza peculiar de muitos problemas em SI, entretanto, pode não ser possível determinar, ou ao menos descrever explicitamente os meios, fins e leis relevantes.

Nesta pesquisa o processo de busca guiado pelo *Design Science Research* apresentou o objetivo de desenvolver um modelo para geração de recomendações de parcerias fundamentado na análise de redes de colaboração científica composta pelos pesquisadores da área de Ciência da Computação.

7. **Comunicação da pesquisa:** *Design Science Research* deve ser representada tanto para o público orientado a tecnologia como para o público gerencial. O público orientado a tecnologia pode receber detalhes suficientes que permitam à construção do artefato usado no contexto organizacional apropriado. Isso permite aos praticantes tirar vantagem dos benefícios oferecidos pelo artefato e permite ainda aos pesquisadores construir uma base de conhecimento cumulativa para posterior extensão e avaliação. *Design Science Research* também é importante para tais audiências entenderem os processos pelos quais o artefato foi construído e avaliado. Isso estabelece ainda a reproduzibilidade do projeto de pesquisa e constrói a base de conhecimento para posteriores extensões.

A comunicação desta pesquisa foi efetivada através da apresentação de artigos em eventos científicos, como WPOS 2014 do CIC/UnB, trilha principal do SBSI 2015 (Estrato Qualis B4) (Costa and Ralha, 2015a) e Workshop de Teses e Dissertações em Sistemas de Informação - WTDSI 2015 (Costa and Ralha, 2015b). Pretende-se ainda publicar o trabalho em periódicos qualificados da área de Ciência da Computação. Também faz parte da comunicação a disponibilização na web do artefato de recomendação através do endereço: <http://scicoll-academicmetrics.rhcloud.com/>.

O paradigma *Design Science Research* é adotado nessa pesquisa, conforme proposto por Hevner et al. (2004), 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 Vaishnavi et al. (2007), *Design Science Research* constitui outra “lente”, contendo um conjunto de técnicas com perspectivas analíticas e sintéticas, para realização de pesquisa em SI.

## 2.2 Recuperação da Informação

Algoritmos e técnica de mineração de dados são usados em grandes volumes de dados para extrair padrões e gerar conhecimento com o objetivo de resolver problemas do mundo real. O processo de descoberta do conhecimento a partir desses dados geralmente envolve as seguintes tarefas: pré-processamento da informação, mineração de dados e pós-processamento. Inúmeros estudos baseados em redes de colaboração tem usado técnicas de recuperação da informação. Muitas outras técnicas como mineração de regras de associação, seleção de características por meio de análise visual, seleção de instância e detecção de anomalias são também muito úteis na recuperação de informação e no gerenciamento do conhecimento, assim como o estudo do comportamento humano usando informações de redes de colaboração (Haider et al., 2015). Quando consideramos não somente o volume dos dados a serem analisados, mas a velocidade com que esses dados devem ser processados e também a variedade das fontes dos dados, estamos diante de um problema que pode ser abordado através de técnicas, como por exemplo *Big Data* (McAfee et al., 2012).

Segundo Baeza-Yates and Ribeiro-Neto (2013), Recuperação da Informação (RI) é uma área abrangente da Ciência da Informação que se concentra principalmente em prover aos usuários o acesso fácil às informações de seu interesse, tratando da representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse.

Ainda segundo Baeza-Yates and Ribeiro-Neto (2013), apesar de sua maturidade, até recentemente a área de RI era vista como uma área de interesse limitada apenas a bibliotecários e a especialistas em informação. Essa visão prevaleceu por muitos anos, apesar da rápida disseminação, entre os usuários de computadores pessoais modernos, de ferramentas de RI para aplicações multimídia e hipertexto. No início dos anos 90 um fato mudou de uma vez por todas essa percepção, o surgimento da *World Wide Web* (Web).

A Web, apresentada em 1989 por Tim Bernes-Lee, tornou-se um repositório da cultura e do conhecimento humano. Seu sucesso baseia-se na concepção de uma interface de usuário padrão que é sempre a mesma, não importando o ambiente computacional usado para executá-la, e que permite que qualquer usuário crie seus próprios documentos. Como resultado, milhões de usuários criaram bilhões de documentos que compõem o maior repositório humano do conhecimento na história. Um consequência imediata é que encontrar informações úteis na Web não é sempre uma tarefa simples e, normalmente, requer a submissão de uma consulta a uma máquina de busca, a qual é a materialização dos conceitos e tecnologias de RI.

### Métricas de Avaliação de RI

Uma vez realizada a recuperação de informações de quaisquer fontes, é necessário definir critérios para mensurar quantitativamente a eficiência da recuperação. Para isso, as métricas precisão (*precision*), revocação (*recall*) e média harmônica (*F<sub>-</sub>measure*) são definidas.

Antes de descrever como a avaliação da RI é realizada, faz-se necessário apresentar o conceito de coleção de referência. Uma coleção de referência é composta por um conjunto



$\mathcal{D}$  de documentos pré-selecionados, um conjunto  $I$  de descrições de necessidade de informação usado para teste e um conjunto de julgamentos de relevância binários associadas a cada par  $[i_m, d_j]$ ,  $i_m \in I$  e  $d_j \in \mathcal{D}$ . O julgamento de relevância tem valor 0 se o documento  $d_j$  não for relevante para a necessidade de informação  $i_m$  e o valor 1 se  $d_j$  for relevante para  $i_m$ .

(Baeza-Yates and Ribeiro-Neto, 2013) fornecem a seguinte descrição para precisão e revocação em um conjunto de documentos. Considere uma requisição de informação  $I$  e seu conjunto  $R$  de documentos relevantes. Seja  $|R|$  o número de documentos nesse conjunto. Suponha que um dado algoritmo de recuperação (que está sendo avaliado) processa a requisição de informação  $I$  e gera um conjunto de resposta  $A$ . Seja  $|A|$  o número de documentos nesse conjunto e  $|R \cap A|$  a interseção dos conjuntos  $R$  e  $A$ , conforme ilustrado na Figura 2.2.

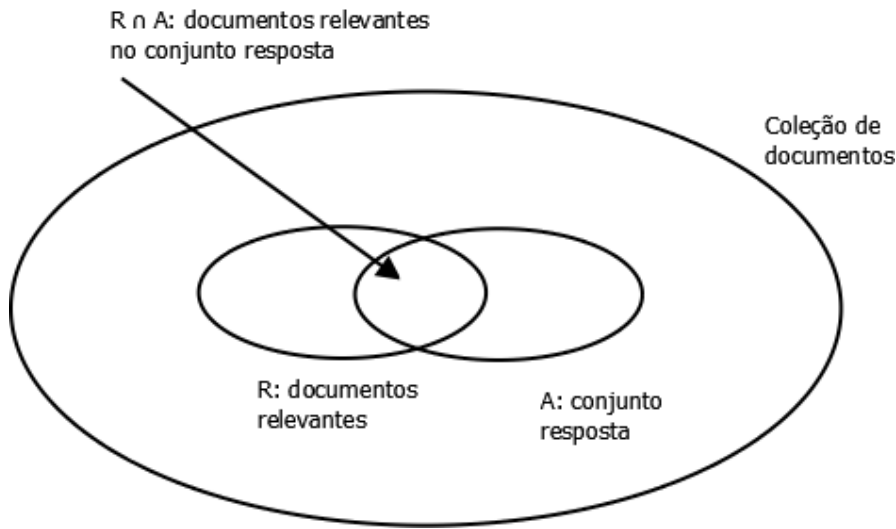


Figura 2.2: Precisão e Revocação em um conjunto de documentos

A métrica *Precision* é usada neste trabalho como a fração das parcerias recomendadas que foram efetivamente realizadas no ano seguinte e expressada através da Equação 2.1. Salienta-se que foi usado o intervalo de tempo de um ano (ano seguinte), porém esse parâmetro pode ser ajustado para um intervalo maior.

$$precision = p = \frac{|R \cap A|}{|A|} \quad (2.1)$$

A métrica de *Recall* é definida como sendo a fração das parcerias que ocorreram na massa de dados de teste no ano seguinte, assim como para o cálculo da *Precision*, e que figuraram também nas recomendações produzidas, conforme Equação 2.2.

$$recall = r = \frac{|R \cap A|}{|R|} \quad (2.2)$$

A métrica de *F<sub>1</sub> measure*, também conhecida como média harmônica, é uma maneira de combinar *Precision* e *Recall* em um único valor e é expressa pela Equação 2.3, onde  $p$

e  $r$  representam *precision* e *recall*, respectivamente e  $\lambda$  consiste em um fator usado para ajustar o peso dos índices *Precision* e *Recall* no cálculo de *F<sub>-measure</sub>*.

$$F_{-measure} = \lambda * \frac{p * r}{p + r} \quad (2.3)$$

*Precision*, *Recall* e *F<sub>-measure</sub>*, conforme definidos, supõem que todos os documentos (recomendações) no conjunto de respostas  $A$  foram examinados (ou vistos). Contudo, o usuário não é normalmente apresentado a todos os documentos do conjunto de resposta  $A$  de uma vez. Em vez disso, os documentos do conjunto  $A$  são primeiro ordenados de acordo com uma função de ranqueamento. O usuário, então, examina essa lista ordenada iniciando pelo documento do topo. Nesse caso, em um sistema interativo, as medidas de *Precision*, *Recall* e *F<sub>-measure</sub>* variam conforme o usuário procede com seu exame do conjunto resposta (Baeza-Yates and Ribeiro-Neto, 2013).

## 2.3 Redes de Colaboração

Segundo Newman (2001), pesquisas tradicionais em redes sociais tem sido conduzidas em estudos de campo. Tipicamente o pesquisador busca por comunidades auto contidas como a comunidade escolar, religiosa ou étnica e constrói a rede de relacionamentos através da entrevista de seus participantes, ou circulando questionários. Um estudo pedirá aos participantes para nomear as pessoas que lhe são próximas, provavelmente classificadas por proximidade. Podem também, opcionalmente, solicitar mais informações sobre essas pessoas e sobre a natureza do relacionamento entre elas.

Estudos desse tipo têm revelado muito sobre a estrutura das comunidades, porém sofrem de dois problemas sérios que as tornam pobres fontes de informação para o tipo de abordagem para a análise de rede. No primeiro, as informações que eles retornam não são numerosas. Coletar e compilar as informações desses estudos é um árduo processo e a maioria dos conjuntos de dados não possuem mais que dezenas ou centenas de atores. É raro um estudo que exceda 1.000 atores. Isto dificulta a precisão dos cálculos estatísticos de muitos resultados. No segundo, eles contêm erros significantes e não controlados como resultado da natureza subjetiva das respostas dos participantes. O que um participante considera ser um relacionamento de amizade, por exemplo, pode ser completamente diferente do que um outro participante responde. Em estudos com crianças em idade escolar, por exemplo, pode-se constatar que algumas crianças considerarão ter amizade com qualquer um da centena de colegas de escola, enquanto outros considerarão um ou dois amigos. Claramente esses participantes estão empregando definições diferentes de amizade. Uma fonte de informações mais promissora é a rede de “afiliação”.

Uma rede de afiliação é uma rede de atores agrupados por algum tipo de relacionamento, como clubes, times ou organizações. Exemplos que têm sido estudados no passado incluem presidentes de companhias e os clubes que eles frequentam (Galaskiewicz and Marsden, 1978), diretores e os conselhos diretivos dos quais eles fazem parte MARIOLIS (1975); Davis and Greve (1997), mulheres e os eventos sociais que elas frequentam (Davis et al., 2009).

Dados em redes de afiliações tendem a ser mais confiáveis do que em outras redes sociais, dado que a filiação a um grupo frequentemente pode ser determinada com uma precisão não disponível quando consideramos algum tipo de amizade. Grandes redes

podem ser construídas dessa forma, uma vez que em muitos casos a inclusão de um ator em um grupo pode ser verificada com a lista de membros do grupo, fazendo com que entrevistas e questionários se tornem desnecessários. Uma rede de atores de filmes, por exemplo, e os filmes nos quais eles atuaram pode ser compilada usando recursos da Internet Movie Database<sup>1</sup>, contendo quase meio milhão de atores, uma amostra melhor para se analisar estatísticas que a maioria das redes sociais, embora não esteja claro se essa rede em particular tenha algum interesse social real.

Assim como as rede de afiliações, redes de colaboração científica são redes sociais que se diferenciam pela característica de seus relacionamentos que representam algum tipo de interação científica. Segundo Freire and Figueiredo (2011), 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. Os vértices de uma rede social costumam fazer referência a pessoas, porém podem representar ainda organizações ou eventos. 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 como uma propriedade de uma aresta no grafo. Um grafo dirigido com dois vértices interligados por uma aresta de peso 300 pode ser visualizado na Figura 2.3.

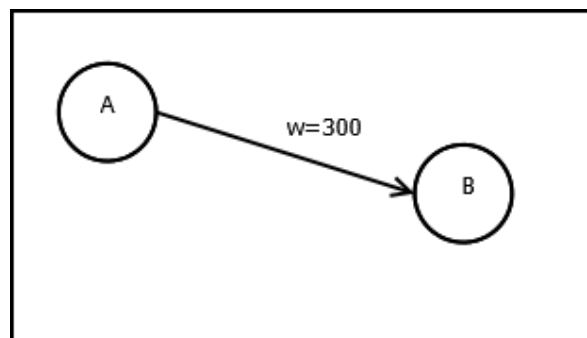


Figura 2.3: Grafo direcionado com dois vértices A e B e uma aresta direcionada de A para B, com peso  $w = 300$ .

Ströele et al. (2013) conceitua genericamente uma rede como sendo 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.

Outra definição para redes de colaboração científicas pode ser encontrada em Barabási et al. (2002) ao descrever o exemplo de um dos matemáticos mais produtivos de todos os tempos, Paul Erdős. Durante sua carreira, Erdős produziu cerca de 1.400 artigos com mais

<sup>1</sup>Disponível em: <http://www.imdb.com/>

de 500 coautores. Essa grande produtividade inspirou o conceito do número de Erdős, que é definido como 1 para seus coautores, 2 para os coautores desses coautores e assim sucessivamente. O alto grau de interconexão da comunidade científica é refletida pela conjectura de que todos os autores matemáticos, assim como muitos físicos e economistas possuem pequenos números de Erdős<sup>2</sup>. Além do interesse imediato em “cientometria”, as redes de colaboração são de interesse geral por auxiliarem no entendimento da dinâmica e da topologia de redes complexas, como as representadas pelas grandes redes sociais existentes.

Redes de colaboração têm sido muito estudadas nas ciências sociais (Wasserman and Faust, 1994; Kochen, 1989). Uma característica geral desses estudos é que eles são restritos a pequenos sistemas, e frequentemente visualizam essas redes como grafos estáticos, cujos vértices representam indivíduos e as arestas representam interações sociais passíveis de quantificação.

Por outro lado, abordagens recentes usam metodologias baseadas em estatísticas direcionadas a grandes redes, buscando universalidades tanto na topologia da rede quanto na dinâmica que governa sua evolução. Esses resultados teóricos e práticos combinados abriram os caminhos para o direcionamento de pesquisas em uma vasta área de abrangência desde a Ciência da Computação até a Biologia. Três importantes resultados parecem ter sido cristalizados como resultado dessas pesquisas: (i) a maior parte das redes possuem um propriedade chamada “pequeno mundo”, que significa que a separação média entre vértices é pequena, ou seja, é possível encontrar um pequeno caminho entre a maioria dos pares de vértices; (ii) redes reais demonstram possuir um grau de agrupamento de vértices maior que o apresentado por redes aleatórias; (iii) tem-se demonstrado que o grau de distribuição contém importantes informações sobre natureza da rede como muitas redes extensas que obedecem a distribuição de potência livres de escala.

O fato do grau de distribuição conter importantes informações sobre a natureza da rede se dá como consequência de dois mecanismos genéricos: (i) a expansão contínua da rede pela adição de novos vértices; e (ii) a ligação de novos vértices preferencialmente a outros vértices bem conectados (Barabási and Albert, 1999), cuja definição deu origem ao modelo Barabási-Albert para geração de grafos que representem redes sociais. Um exemplo pode ser observado pela Figura 2.4 que apresenta a distribuição de relacionamentos de autoria por autor, destacando que poucos autores possuem uma quantidade bem elevada de publicações (relacionamento) enquanto a maioria dos autores colabora com poucos relacionamentos. A Figura 2.4 foi gerada a partir dos dados constantes no banco de dados do estudo de caso implementado para essa pesquisa.

Para ilustrar a conexão entre vértices como parte de uma rede de colaboração, a Figura 2.5 representa essa rede como um grafo com centro em um pesquisador. Nele, os vértices azuis representam as publicações realizadas pelos pesquisadores (utilizando o ISSN do veículo de publicação) que são representados pelos vértices roxos. Os relacionamentos entre os vértices são implementados pelas arestas com rótulo *AUTHORING*.

---

<sup>2</sup>Projeto Número de Erdős disponível em: <http://wwwp.oakland.edu/enp/>

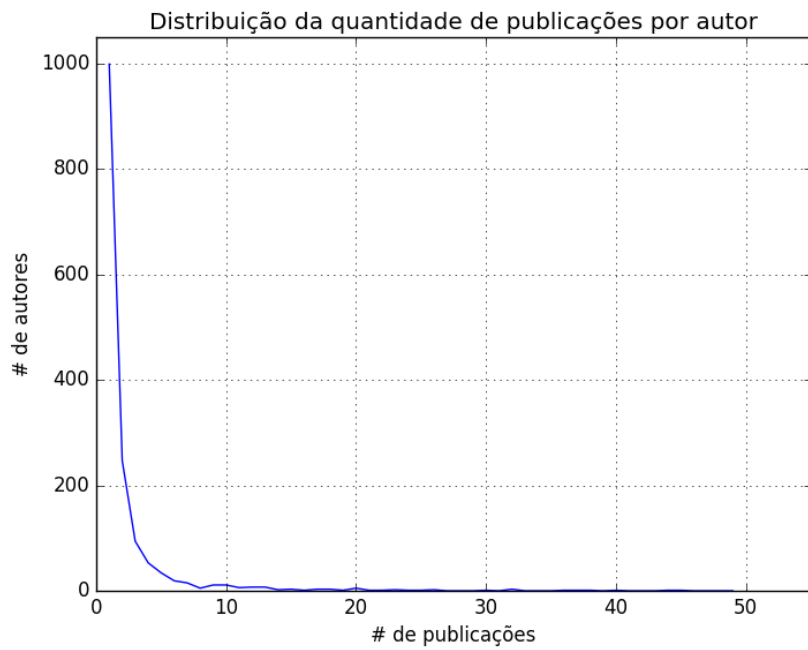


Figura 2.4: Distribuição da quantidade de publicações por autor

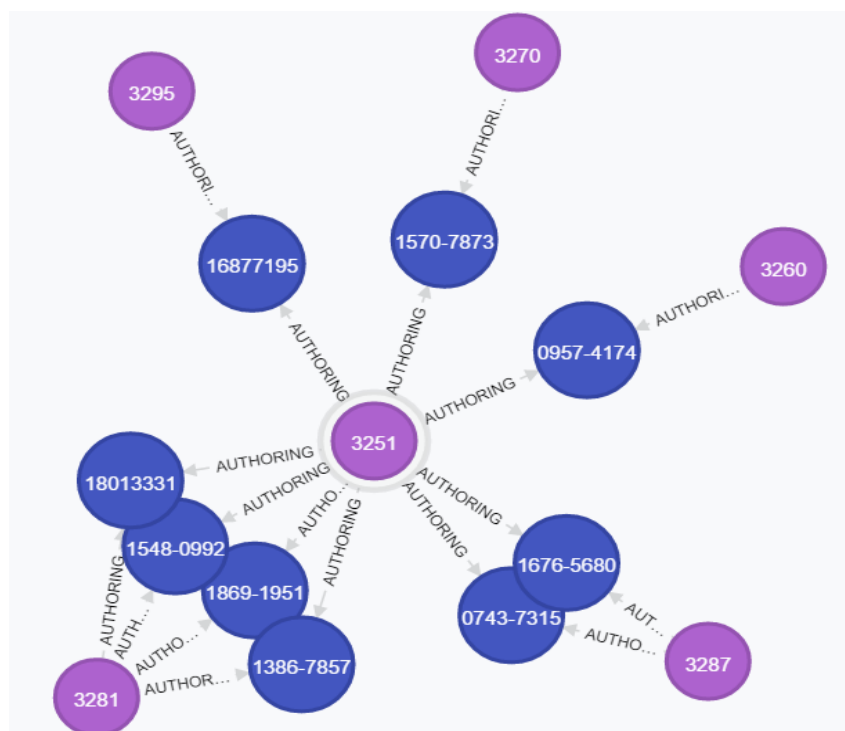


Figura 2.5: Rede de colaboração centrada um pesquisador

## 2.4 Sistemas de Recomendação

Segundo Resnick and Varian (1997), fazer escolhas com pouca ou nenhuma experiência pessoal é uma tarefa corriqueira do ser humano. Para tanto, são utilizadas recomendações, seja na forma de críticas de cinema e livro ou até mesmo recomendações de seres humanos através de cartas. Os Sistemas de Recomendação (SR) são portanto uma ferramenta para melhorar um processo social natural, que é fazer escolhas.

Antes de apresentar os conceitos de SR, consideremos inicialmente os métodos de predição de linha de base de forma a melhor contextualizar esses sistemas. Os métodos de predição de linha de base são úteis para o estabelecimento de linhas de base não personalizadas contra as quais algoritmos de personalização (sistemas de recomendação) podem ser comparados. Métodos de definição de linha de base que não dependem da avaliação do usuário para um item também podem ser úteis para fornecer predições para novos usuários, como uma forma de contornar o problema de “partida fria” dos SR, esse problema consiste na falta de informações que o modelo pode usar para efetuar as predições, muito comum na inicialização do modelo. A linha de base predita para o usuário  $u$  e para o item  $i$  é  $b_{u,i}$ . A linha de base mais simples de predizer para a avaliação de um item é a média das avaliações fornecidas para este item, ou seja  $b_{u,i} = \mu$  (onde  $\mu$  é a média geral de avaliações). A linha de base pode ser melhorada quando leva em consideração a média das avaliações fornecidas por um usuário  $\bar{r}_u$  ou para um item  $\bar{r}_i$  e combina-se essa métrica conforme a Equação 2.4:

$$b_{u,i} = \mu + b_u + b_i \quad (2.4)$$

onde  $b_u$  e  $b_i$  são preditores da linha de base de usuário e do item, respectivamente. Esses preditores podem ser definidos conforme as Equações 2.5 e 2.6, respectivamente.

$$b_u = \frac{1}{|I_u|} \sum_{i \in I_u} (r_{u,i} - \mu) \quad (2.5)$$

$$b_i = \frac{1}{|U_i|} \sum_{u \in U_i} (r_{u,i} - b_u - \mu) \quad (2.6)$$

As Equações 2.5 e 2.6 usam o conjunto de itens  $I_u$  recomendados para o usuário  $u$  e o conjunto de usuários  $U_i$  os quais receberam recomendações para o item  $i$ , respectivamente. Dessa forma, para um item ou usuário novo sua linha de base pode ser considerada zero, assumindo que essa é a média das avaliações para o item ou para a identificação do usuário (Ekstrand et al., 2011).

No contexto dos SR, a tarefa de recomendar também é conhecida como filtragem, uma vez que consiste na busca de itens relevantes entre todos os itens disponíveis.

SR se diferenciam quanto a forma de analisar os dados com o objetivo de desenvolver a noção de afinidade entre usuários e itens, a qual pode ser usada para identificar o casamento de pares usuário-item. Existem duas formas de filtragem mais difundidas em SR: (i) a filtragem colaborativa, que analisa interações históricas entre usuários sobre um item específico, ou seja, ao interagir com um item, um usuário registra sua afinidade por esse item e essa afinidade pode ser considerada ao se sugerir itens a outro usuário similar ao primeiro; e (ii) 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 de filtragem (Melville and Sindhvani, 2010).

Segundo Goldberg et al. (1992), ao definir um dos primeiros sistemas de recomendação, o *Tapestry*, um SR baseado em filtragem colaborativa é um mecanismo mais eficiente pois seu trabalho envolve intervenção humana no processo. Para exemplificar, Goldberg et al. (1992) afirmam que algum tempo após a chegada de um documento, como um email, um humano pode lê-lo e decidir que é interessante, adicionando esse documento a sua lista de recomendações. Logo percebe-se que é necessário um mecanismo constante de leitura e análise de documentos. Ekstrand et al. (2011) definem matematicamente um sistema de recomendação baseado em filtragem colaborativa através da Equação 2.7, na qual a recomendação é gerada de um item  $i$  para um usuário  $u$ , usando  $s$ , uma função de similaridade, para encontrar  $N \subseteq U$ , o conjunto de vizinhos de  $u$ .

$$p_{u,i} = \bar{r}_u + \frac{\sum_{u' \in N} s(u, u')(r_{u',i} - \bar{r}_{u'})}{\sum_{u' \in N} |s(u, u')|} \quad (2.7)$$

Conforme Ströele et al. (2013), a recomendação de parcerias, ou predição de *link* no contexto de um grafo social, pode ser realizada de duas formas diferentes, porém complementares. A primeira examina um sociograma desbalanceado de estruturas sociais buscando transformá-lo em uma estrutura social balanceada. Como exemplo, utilizando filtragem colaborativa, se duas pessoas possuem muitos amigos em comum então essas duas pessoas devem se conhecer. A segunda abordagem analisa o conteúdo da comunicação entre os indivíduos, buscando por indivíduos que compartilham o mesmo interesse. Como exemplo, utilizando filtragem colaborativa, se dois pesquisadores publicam sobre um mesmo assunto então eles poderiam interagir, nesse caso o método mais adequado para o SR é a filtragem baseada em conteúdo.

Segundo Melville and Sindhvani (2010), o objetivo de um SR é gerar recomendações significativas a um conjunto de usuários para itens ou produtos que possam interessar. Sugestões de livros na Amazon, amigos no Facebook ou filmes no Netflix são exemplos reais do uso de SR que utilizamos no dia-a-dia. O projeto de um SR depende do domínio de características particulares dos dados disponíveis. Tais dados podem armazenar a relevância de um determinado item para o usuário, por exemplo, 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.

## Similaridade

Um conceito que merece destaque na definição de um sistema de recomendação que usa filtragem colaborativa é similaridade. Segundo o Dicionário Português<sup>3</sup>, a definição de similar é: que possuem o mesmo teor, que se assemelham ou se equivalem, que são semelhantes. Formalmente, é preciso usar uma função matemática para definir a semelhança entre objetos. Ekstrand et al. (2011) citam quatro funções de similaridade disponíveis na literatura, embora outros métodos também sejam utilizados:

- **Índice de correlação de Pearson:** Esse método computa a correlação de Pearson  $s$  entre as avaliações em comum de dois usuários  $u$  e  $v$ . A correlação é calculada

---

<sup>3</sup>Disponível em: <http://dicionariportugues.org/>

pela Equação 2.8, onde  $I$  representa o conjunto de itens avaliados por cada usuário e  $r_{v,i}$  a avaliação dada pelo usuário  $v$  para o item  $i$ .

$$s(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \bar{r}_v)^2}} \quad (2.8)$$

A correlação de Pearson calcula altos índices de similaridade entre usuários com poucas avaliações em comum. Isto pode ser minimizado adotando-se um limiar para o número de itens co-avaliados necessários para a convergência de avaliações (correlação de 1) e colocando a similaridade em escala quando o número de itens co-avaliados está abaixo desse limiar (Herlocker et al., 1999, 2002).

- **Correlação de Pearson Restrita:** Usando uma escala fixa de 7 pontos e usando o ponto 4 como valor neutro para avaliação normalizada  $r_z$ , é possível relacionar as avaliações “gostei” e “não gostei” com essa escala, diferente do cálculo padrão da correlação de Pearson que usa uma escala livre. Isso motivou Shardanand and Maes (1995) a propor a correlação de Pearson restrita através da Equação 2.9, na qual os termos são os mesmo usados pela Equação 2.8 porém com a mudança das médias  $\bar{r}_u$  e  $\bar{r}_v$  pela avaliação na escala de 7 pontos  $r_z$ .

$$s(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - r_z)(r_{v,i} - r_z)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - r_z)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - r_z)^2}} \quad (2.9)$$

- **Correlação de Spearman:** Para a correlação de Spearman, os itens que um usuário avaliou são classificados pela ordem da avaliação, ou seja, o item melhor avaliado aparece em primeiro na classificação. Itens com a mesma avaliação são classificados pela média de suas avaliações. O cálculo é o mesmo que a correlação de Pearson, com a ordenação dos itens no lugar apenas da classificação. Uma vantagem da correlação de Spearman sobre a correlação de Pearson é que a primeira é menos sensível a valores muito distantes do esperado (Herlocker et al., 1999, 2002).
- **Similaridade de Cosseno:** Esse modelo usa uma abordagem diferente das anteriores, utilizando uma abordagem baseada no espaço vetorial da álgebra linear ao contrário dos anteriores que usam abordagem estatística. Neste modelo os usuários são representados como vetores  $|I|$ -dimensionais e a similaridade é medida pela distância dos cossenos entre dois vetores de avaliação. O cálculo do coeficiente pode ser realizado através do produto vetorial entre  $R_u$  e  $R_v$  (vetores de avaliação), dividido pela distância euclidiana de seus módulos (Equação 2.10). Avaliações desconhecidas, ou seja, para itens nunca avaliados, são consideradas como zero.

$$s(u, v) = \frac{R_u \cdot R_v}{\|R_u\| * \|R_v\|} = \frac{\sum_i R_{u,i} R_{v,i}}{\sqrt{\sum_i R_{u,i}^2} \sqrt{\sum_i R_{v,i}^2}} \quad (2.10)$$

Existem ainda os métodos de similaridade que consideram os objetos como vetores, é o caso da distância de *Hamming* a qual, dados dois vetores de mesmo tamanho com valores binários será igual a quantidade de elementos divergentes nesses vetores, conforme



definido por Shimodaira (2014). E a distância de *Levenshtein* é definida, segundo Navarro (2001) como o menor número de operações de inserção, remoção e troca para tornar duas cadeias de caracteres iguais. Matematicamente, a distância de *Levenshtein* é definida para duas cadeias de caracteres  $a$  e  $b$  de tamanhos  $|a|$  e  $|b|$ , respectivamente, conforme a Equação 2.11.

$$lev_{a,b}(i, j) = \begin{cases} \max(i, j), & \text{if } \min(i, j) = 0 \\ \min(i, j) \begin{cases} lev_{a,b}(i-1, j) + 1, \\ lev_{a,b}(i, j-1) + 1, \\ lev_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)}, \end{cases} & \text{caso contrário.} \end{cases} \quad (2.11)$$

Nessa pesquisa foi adotada uma abordagem híbrida combinando a filtragem colaborativa baseada em vizinhança (página 18, último parágrafo), através da similaridade entre os pesquisadores por meio das áreas de pesquisa. Também foi adotada a filtragem baseada em conteúdo usando os dados históricos das publicações em periódicos. Neste sentido, representa-se cada pesquisador como um vetor de áreas de pesquisa e emprega-se a distância de *Hamming* conforme definido por Shimodaira (2014). Desta forma, quanto maior a quantidade de áreas de pesquisa em comum entre dois pesquisadores, mais similares ele são. A distância de *Levenshtein* foi utilizada para desambiguação dos nomes dos pesquisadores, conforme será apresentado na Seção 4.1.

Os fundamentos apresentados no presente capítulo são explorados nesta pesquisa tanto para o desenvolvimento do modelo de recomendação como para a avaliação das recomendações produzidas. Esses fundamentos também são explorados nos trabalhos correlatos apresentados no Capítulo 3.

# Capítulo 3

## Trabalhos Correlatos

Para a realização da pesquisa do estado da arte foi utilizado o método de revisão *quasi*-sistemática, conforme definido por Kitchenham (2004), apoiado pela ferramenta StArt (Zamboni et al., 2010). Marshall et al. (2014) realizaram uma análise comparativa de várias ferramentas para realização de revisões sistemáticas. As ferramentas analisadas foram SLuRp, StArt, SLR-Tool e SLRTOOL. Nessa análise, o StArt ficou em segundo colocado com uma pontuação de 55.3%, tendo sido escolhida nesta pesquisa por possuir o processo de instalação mais simples, entre as ferramentas, bem como possibilitar o acompanhamento de todas as etapas de Revisão Sistemática. Embora não disponha de suporte a edição colaborativa.

As seções desse capítulo estão assim organizadas: na Seção 3.1 são apresentados os conceitos de Revisão *Quasi*-sistemática, bem como a definição e a execução do protocolo aplicado; e na Seção 3.2 são apresentados os trabalhos que diretamente serviram de base para esta pesquisa.

### 3.1 Revisão *Quasi*-sistemática

Uma revisão sistemática é um meio de avaliar e interpretar toda pesquisa disponível e relevante a uma questão de pesquisa em particular, área de pesquisa, ou fenômeno de interesse. Revisões sistemáticas visam apresentar uma clara avaliação de um tópico de pesquisa através do uso de uma metodologia confiável, rigorosa e auditável.

Segundo Biolchini et al. (2005), a revisão sistemática consiste em uma metodologia científica específica que vai um passo além de uma simples revisão da literatura. Ela busca integrar pesquisas empíricas para criar generalizações. Essa tarefa de integração envolve objetivos específicos, os quais permitem ao pesquisador analisar criticamente os dados coletados, de forma a resolver conflitos detectados na literatura utilizada e identificar pendências para o planejamento de futuras investigações. Devido a essa característica particular, a revisão sistemática não é considerada apenas uma fase de uma pesquisa, papel desempenhado pela revisão usual da literatura, mas uma forma de se estruturar a pesquisa. Salienta-se que uma revisão *quasi*-sistemática é uma revisão preliminar sobre determinado assunto de pesquisa, enquanto uma revisão sistemática é uma revisão mais completa que pode ser realizada com base em uma revisão *quasi*-sistemática ou sistemática.

A condução de uma revisão *quasi*-sistemática é realizada em três fases: planejamento, execução e sumarização ou análise de resultados. Na fase de planejamento, os objetivos

da pesquisa são definidos assim como a forma pela qual a revisão será executada. Esses passos incluem a formulação da questão de pesquisa e a definição dos métodos de seleção de fontes de pesquisa e de estudos. Durante a fase de execução da revisão, a pesquisa nas fontes escolhidas deve ser executada e os estudos obtidos devem ser avaliados de acordo com os critérios estabelecidos. Por fim as informações relevantes para a questão de pesquisa devem ser selecionadas a partir dos estudos selecionados. Após a execução da revisão sistemática, os resultados devem ser sumarizados e analisados usando algum método estatístico definido durante a fase de planejamento.

O protocolo que foi utilizado para a condução da revisão *quasi*-sistemática nesta pesquisa, durante o período de janeiro a maio de 2016, conforme descrito na sequência.

- **Questão de pesquisa** – É possível definir e implementar um artefato para análise de redes de colaboração científica que realize recomendações de parcerias meritosas para pesquisadores?
- **População** – Pesquisas sobre redes de colaboração científica, análise de rede sociais e sistemas de recomendação.
- **Intervenção** – Abordagens para análise da dinâmica de redes de colaboração científica analisando-as quanto à densidade de relacionamentos, métricas de centralidade e agrupamentos de vértices; Sistemas de recomendação que adotem técnica híbridas de filtragem de conteúdo e filtragem colaborativa.
- **Controle**
  1. Victor Ströele and Geraldo Zimbrão and Jano M. Souza. Group and link analysis of multirelational scientific social networks. *J. Syst. Softw.*, 86(7):1819-1830, July 2013. ISSN 0164-1212.
  2. Waleej Haider and Nouman M. Durrani and Shardha Nand and Nadeem K. Khan and Muhammad A. Abbasi. Towards knowledge creation and management model over online social networks. In: *The 2<sup>nd</sup> World Symposium on Web Applications and Networking (WSWAN)*, 2015.
  3. Edeilson M. Silva and Ricardo A. Costa and Mario Godoy Neto and 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. In: *Simpósio Brasileiro de Sistemas de Informação (SBSI)*, 2009.
- **Resultados** – Avaliar diferentes métodos de análise de rede de colaboração para selecionar os que serão aplicados no sistema de recomendação;
- **Aplicação** – Ajudar os pesquisadores a visualizar a rede de colaboração científica de um grupo e considerar as recomendações de parceria em pesquisas científicas.

Os repositórios digitais de publicações que incluem pesquisas na área de Ciência da Computação usados para as consultas foram o IEEE Explorer<sup>1</sup>, Scopus<sup>2</sup> e Engineering

---

<sup>1</sup>Disponível em: <http://ieeexplore.ieee.org/>

<sup>2</sup>Disponível em: <https://www.scopus.com/>

Village<sup>3</sup>. O quantitativo de publicações retornadas pelos três repositórios foi de 91 (noventa e uma) publicações, porém 4 (quatro) eram publicações duplicadas e 6 (seis) foram rejeitadas ainda na fase de seleção por não terem foco em redes de colaboração. Esse quantitativo está demonstrado na Figura 3.1 que exibe ainda o percentual de colaboração de cada fonte de publicação na quantidade de publicações consideradas.

As consultas realizadas em cada repositório digital foram as seguintes:

- **Engineering Village:** “scientific collaboration network” OR “social graph analysis” OR “social network” OR “system dynamics” OR “network dynamics” OR “knowledge flow” OR “graph database” OR “academic network” AND (lattes OR dblp OR bdbcomp)
- **IEEE:** (“scientific network” OR “social graph” OR “social network” OR “network dynamics” OR “knowledge flow” OR “graph database” OR “academic network” OR lattes OR dblp OR bdbcomp) AND (method OR framework OR analysis) AND (academic OR scientific)
- **Scopus:** TITLE-ABS-KEY ( “scientific collaboration network” OR “social graph analysis” OR “social network” OR “system dynamics” OR “network dynamics” OR “knowledge flow” OR “graph database” OR “academic network” AND ( lattes OR dblp OR bdbcomp ) )

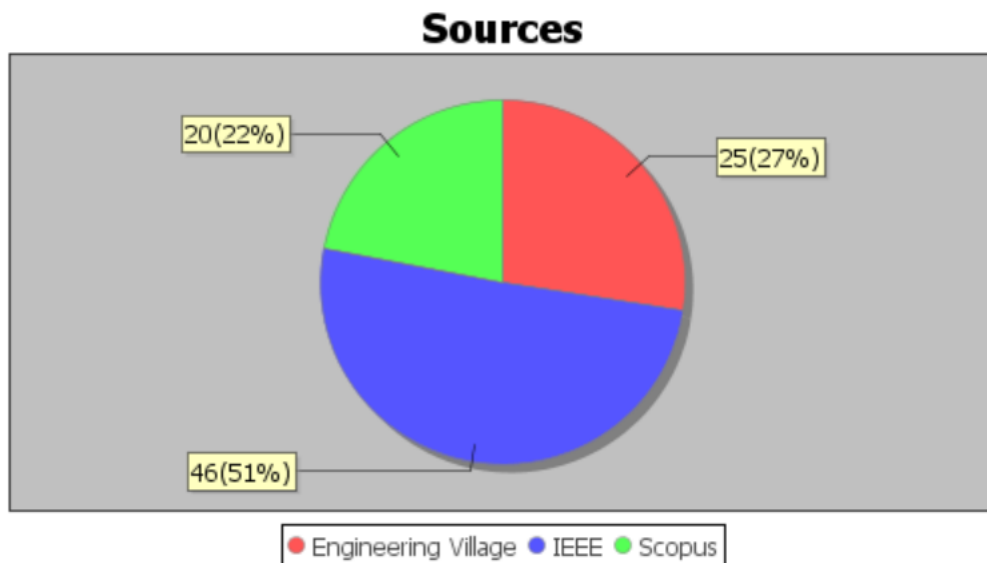


Figura 3.1: Repositórios de publicações usados.

A fase de extração das publicações resultou na aceitação de 74 (setenta e quatro) publicações e 7 (sete) rejeições, conforme Figura 3.2. As Figuras 3.3 e 3.4 exibem os critérios que foram utilizados para aceitar e rejeitar, respectivamente, as publicações selecionadas.

Os critérios usados para aceitação das publicações foram: Foco em redes sociais, Exibir resultados conclusivos e Foco na dinâmica de adaptação dos sistemas. Já os critérios

<sup>3</sup>Disponível em: <https://www.engineeringvillage.com/>

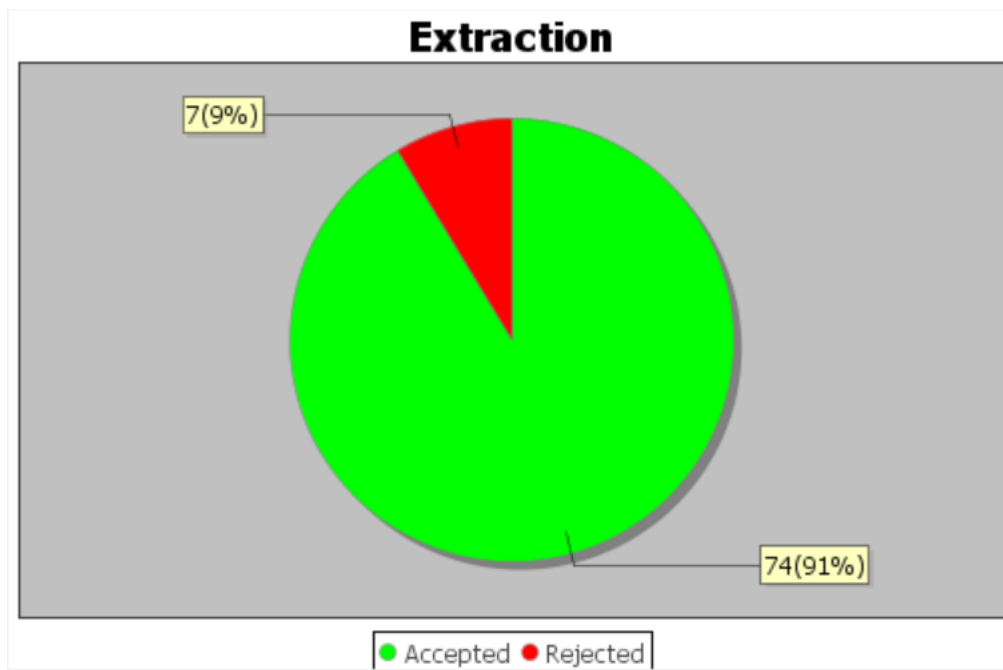


Figura 3.2: Resultado da fase de extração das publicações.

de rejeição foram: Indisponibilidade do documento completo, Documento repetido, Publicação na forma de editorial, resumo ou poster, Documento não escrito em inglês ou português e Foco em redes não-sociais.

O critério Foco em redes não-sociais foi adotado como majoritário para rejeitar as publicações, na fase de extração da revisão *quasi*-sistemática, conforme mostrado na barra cinza da Figura 3.4.

## 3.2 Principais Artigos

Conforme exposto na Seção 3.1 de revisão *quasi*-sistemática, foram selecionados 74 artigos de acordo com os critérios de aceitação estabelecidos. Desse total, quatro artigos foram considerados inspiradores para o presente trabalho, os quais serão apresentados nesta seção.

Ströele et al. (2011a) apresentaram um estudo de identificação de agrupamentos de pesquisadores com o objetivo de analisar o fluxo de informação intra e inter grupos de pesquisa. O trabalho utilizou os currículos dos pesquisadores disponibilizados na plataforma Lattes, vinculados aos programas de Pós-Graduação na área de Ciência da Computação que apresentavam em 2011 nota 6 ou 7 na avaliação da CAPES. Através do método usado, foi possível comprovar a existência de grupos de pesquisa interdisciplinares bem como as áreas de pesquisa que estão interligadas mais fortemente entre si, como é o caso de Banco de Dados, Engenharia de Software e SI. A técnica baseada em árvore geradora mínima de Dijkstra (1959) foi empregada para identificação de agrupamentos e serviu de referência para nosso estudo, como primeiro passo para identificação dos centróides de cada área de pesquisa. No presente trabalho, um centróide de uma área é o pesquisador que mais se relaciona com outros pesquisadores que atuam na mesma sub-área da computação.

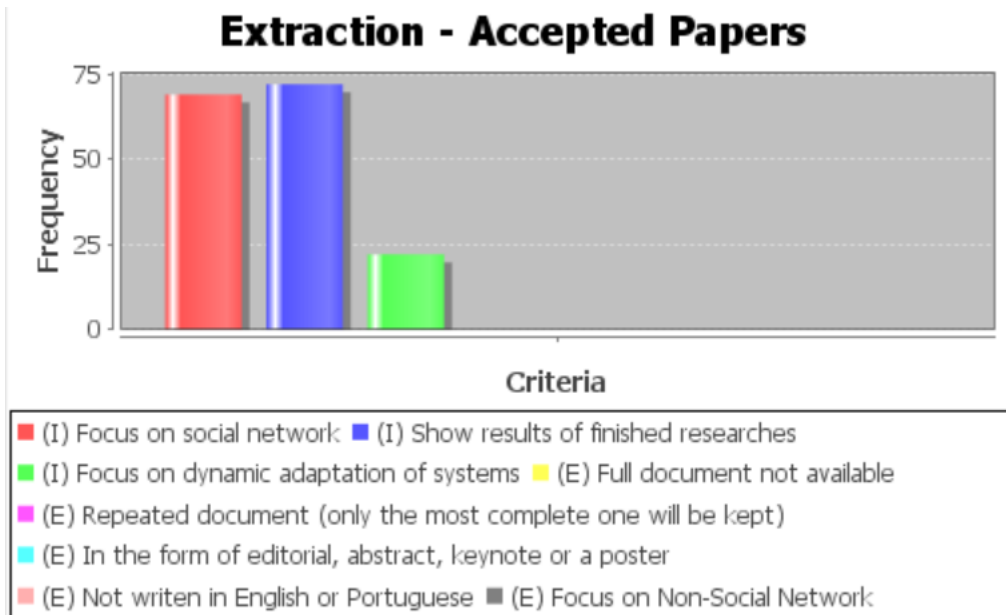


Figura 3.3: Frequência dos critérios usados pelas publicações aceitas.

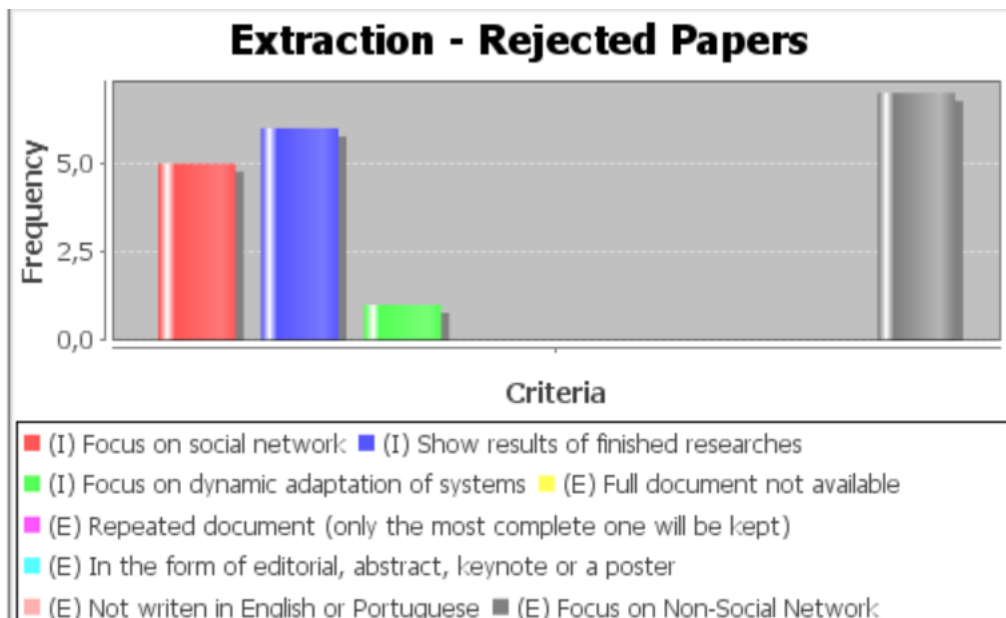


Figura 3.4: Frequência dos critérios usados pelas publicações rejeitadas

Os trabalhos realizados por Ströele et al. (2013) e Ströele et al. (2011b) introduzem várias métricas, como a centralidade de Katz (1953) e o cálculo de clusterização de Wasserman and Faust (1994) usando fluxo máximo de Ford and Fulkerson (1956). O objetivo dessas métricas é auxiliar na análise de redes de colaboração científicas buscando identificar grupos de pesquisadores com interesses em comum no desenvolvimento de suas pesquisas. Para validar essas métricas, foi gerada uma rede de colaboração científica multi-camada, onde cada camada representa um tipo de relacionamento Coautoria, Participação em projeto, Orientação e Produção técnica. Através dessa rede foram realizadas análises como a aplicação de algoritmos de clusterização para identificar os grupos de pesquisadores existentes. Nos trabalhos citados foi apresentada a métrica de quantidade de relacionamentos em comum, ponderados temporalmente, que serviu de motivação para o cálculo do fator de recomendação apresentado no presente trabalho, ou seja, relacionamentos mais antigos possuem uma influência menor na rede. As rede de colaboração científica multi-camadas foram objeto de discussões nesta pesquisa, podendo ser implementado através de uma extensão do artefato atual, uma vez que demanda dados do Currículo Lattes dos pesquisadores para as informações de publicação em eventos científicos, orientação, projetos de pesquisa e produção técnica.

O trabalho realizado por Luo et al. (2014) visou a construção de um sistema de recomendação denominado Hete-CF, baseado em filtragem colaborativa usando uma abordagem heterogênea de relacionamentos entre autores da base DBLP. Os autores citam como abordagens do estado-da-arte o Trust-MF (Yang et al., 2013) e o Hete-MF (Yu et al., 2013). Como métodos de avaliação dos resultados foram utilizados o erro médio absoluto (Mean Absolute Error – MAE) e a raiz quadrada do erro médio (Root Mean Square Error – RMSE). Os métodos de avaliação apresentados serão explorados como trabalhos futuros nesta pesquisa, com a finalidade de comparar os resultados das métricas atualmente aplicadas.

Por fim, consideramos o trabalho desenvolvido por Digiampietri et al. (2015), o qual aborda o problema da desambiguação de nomes de autores e propõem uma forma de solucioná-lo, baseada em quatorze características de cada autor agrupadas por características da rede social de coautorias, características extraídas dos nomes, características baseadas na mineração de texto das publicações e características baseadas nos anos de publicações. A solução do problema de desambiguação de nomes é de fundamental importância para a recuperação da produção científica dos pesquisadores e a posterior criação dos relacionamentos entre autores. Uma vez que os relacionamentos considerados nesse trabalho são descobertos através dos nomes dos autores, uma técnica de desambiguação como a apresentada é importante para melhorar o desempenho do artefato. Optou-se por utilizar um método heurístico baseado em características extraídas dos nomes dos autores usando para isso a distância de *Hamming*, descrita na Seção 2.4, conjugada com características da rede de colaboração entre os pesquisadores como vizinhos em comuns, conforme será apresentado no Capítulo 4.

# Capítulo 4

## Solução Proposta

Este capítulo apresenta o modelo conceitual do sistema de recomendação, descrevendo o fluxo da informação necessária para geração e avaliação das recomendações e também como os dados foram modelados na solução. Na Seção 4.1 uma visão geral da solução é apresentada, na Seção 4.2 o modelo de dados adotado nesta pesquisa é apresentado e na Seção 4.3 as tarefas que compõem a solução são detalhadas.

### 4.1 Modelo Conceitual

A solução proposta contempla a definição de um modelo de recomendações e também a arquitetura de um sistema de recomendação que implementa o modelo. O *workflow* da solução está modelado como um fluxo de processos, mostrado na Figura 4.1, cujas atividades são as seguintes:

1. O processo é iniciado com os dados brutos da rede de colaboração presentes no repositório DBLP. A partir da lista de pesquisadores iniciais e de seus dados de publicações carregados da DBLP, a rede de colaboração é formada e os relacionamentos dessa rede são formados.
2. O grafo da rede de colaboração é representado matricialmente, permitindo a realização de cálculos analíticos da rede de colaboração como caminho máximo entre autores, grau médio, grau máximo, entre outros.
3. As áreas de atuação de cada pesquisador foram identificadas usando o Sistema de Classificação da ACM e o nome de cada periódico onde os pesquisadores publicaram artigos científicos. O uso do Sistema de Classificação da ACM permite que se possa realizar comparações das áreas de atuação de pesquisadores internacionais, ao contrário da Linguagem de Marcação da Plataforma Lattes descrita por dos Santos Pacheco and Kern (2001) que define áreas de pesquisa em língua portuguesa apenas, o que reduz o escopo da pesquisa.
4. O agrupamento dos pesquisadores por áreas e a primeira avaliação do modelo são realizadas, verificando-se o desempenho da identificação das áreas de pesquisa.
5. A geração das recomendações de parcerias é realizada usando, além das áreas de pesquisa identificadas automaticamente, o ano de publicação em cada área identificada e as parcerias já realizadas. Após a realização dessa tarefa mais uma avaliação



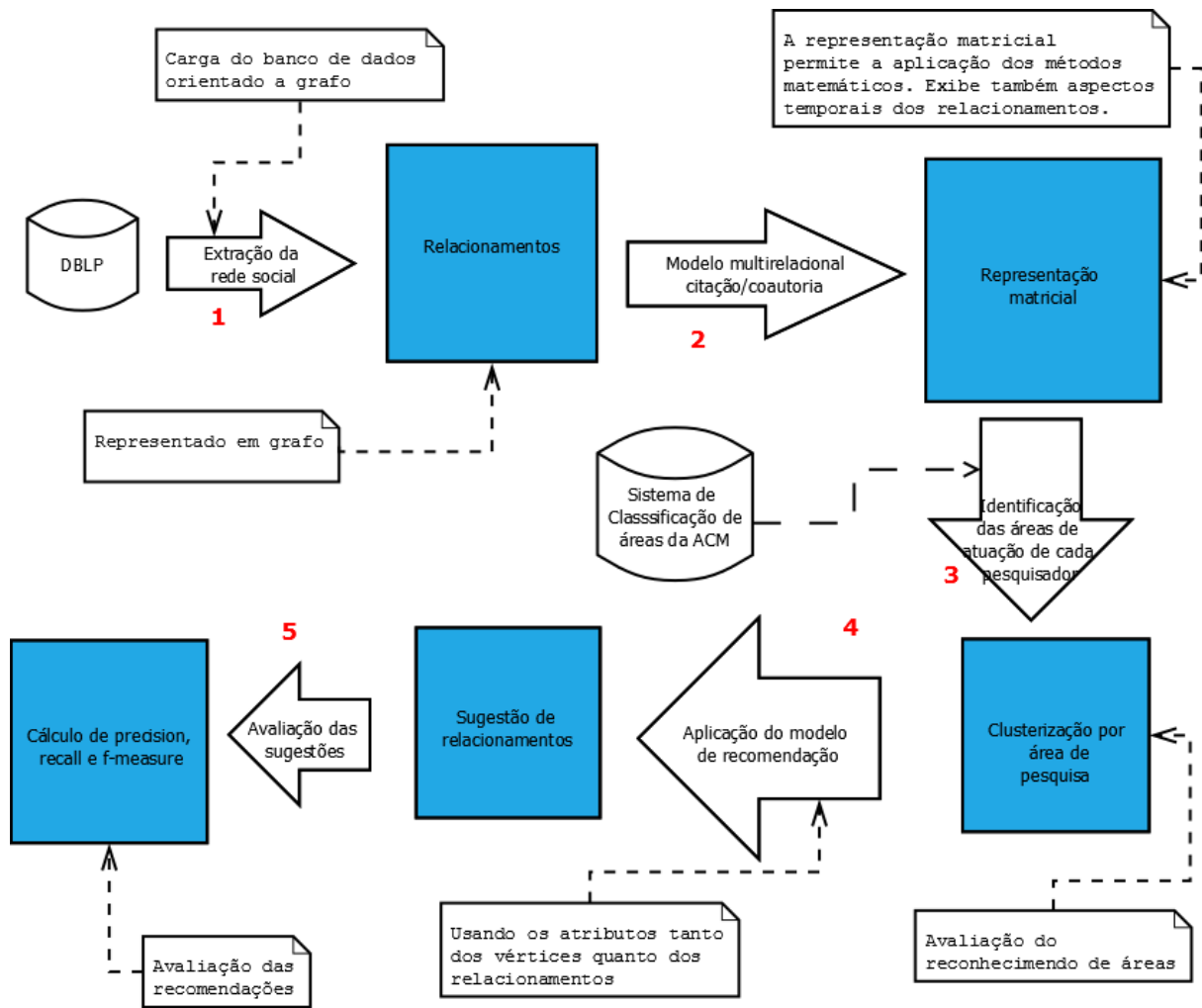


Figura 4.1: *Workflow* da Solução Proposta.

é executada, sendo essa centrada na precisão e na revocação das recomendações realizadas conforme apresentado na Seção 2.2.

Um aspecto que merece especial atenção na solução proposta é quanto ao repositório digital de publicações na área de Ciência da Computação utilizado (DBLP). Existem diversos repositórios de publicações e de currículo de pesquisadores espalhados pela Internet como a plataforma Lattes do CNPq, para currículos de pesquisadores brasileiros, a PubMed, para publicações na área de ciências médicas e a Microsoft Academics, para publicações científicas em diversas áreas. Porém o único repositório que reuniu os requisitos de ser curado e disponibilizar integralmente os dados foi o DBLP<sup>1</sup>, motivo pelo qual este foi escolhido para extração da rede de colaboração a ser trabalhada nessa pesquisa. Um trecho do arquivo XML com o repositório da DBLP pode ser visualizado no Apêndice B.

A rede social científica abordada nesta pesquisa consiste na rede formada pelas relações de coautoria de artigos publicados em periódicos e conferências na área de Computação que estão presentes no repositório de publicações DBLP (totalizando 93,59% conforme a Figura 4.2). Embora outros tipos de publicações como livros, teses, dissertações, entre

<sup>1</sup>Disponível em: <http://dblp.uni-trier.de/xml/>

outros também estivessem presentes na DBLP, optou-se por não usar esses itens pela baixa representatividade, conforme é possível verificar na Figura 4.2. Todavia, é possível a inclusão desses dados com a devida configuração da ferramenta de extração.

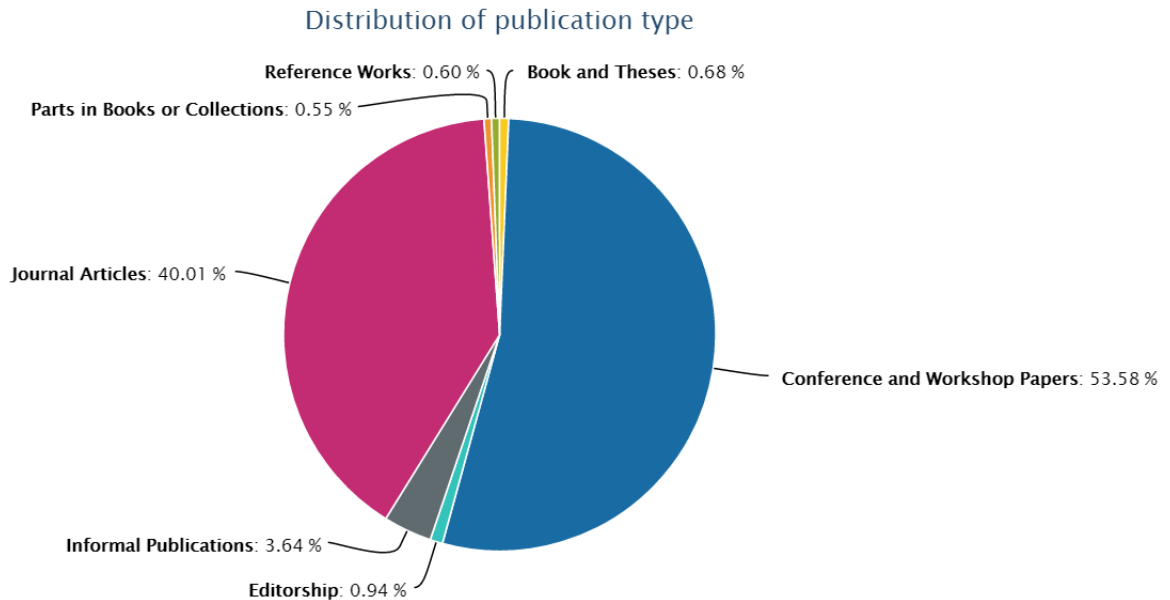


Figura 4.2: Distribuição dos tipos de publicações presentes na DBLP<sup>2</sup>.

Um aspecto que recebeu especial atenção foi a padronização de nomes de áreas de pesquisa. Dessa forma, optou-se por utilizar o Sistema de Classificação de áreas de Computação da ACM<sup>3</sup> para atribuir a cada autor uma ou mais áreas de pesquisa constantes no sistema de classificação da ACM. Com isso, diminui-se a subjetividade na atribuição de áreas aos pesquisadores, permite o relacionamento de áreas entre pesquisadores nacionais e internacionais assim como referenciar as áreas de pesquisa que são informadas explicitamente por alguns periódicos para publicação, como é o caso do periódico *IEEE Transactions on Network Science and Engineering* que declara áreas de interesse incluídas no sistema de classificação da ACM também como *Communication, Networking & Broadcasting, Components, Circuits, Devices & Systems, Computing & Processing* e *Signal Processing & Analysis*.

Embora o Sistema de Classificação da ACM seja hierarquizado, conforme mostrado pela Figura 4.3, que exhibe um trecho da representação do sistema de classificação no formato de árvore, optou-se por adotar até o terceiro nível do sistema de classificação como área de pesquisa englobando todas as subáreas que estão abaixo desse nível. Um exemplo dessa abstração são as áreas *Security in hardware* e *Systems security* que são consideradas como *Security and privacy*, uma vez que as duas primeiras são especializações da terceira, segundo o sistema de classificação da ACM. Uma lista de todas as áreas de pesquisa identificadas pode ser visualizada no Apêndice C.

Outro aspecto que recebe atenção é a desambiguação de nomes de autores em citações nas publicações. Essa multiplicidade de nomes gera dificuldade na comparação dos nomes

<sup>2</sup>Disponível em: <http://dblp.uni-trier.de/statistics/distributionofpublicationtype.html>

<sup>3</sup><http://www.acm.org/about/class/2012>

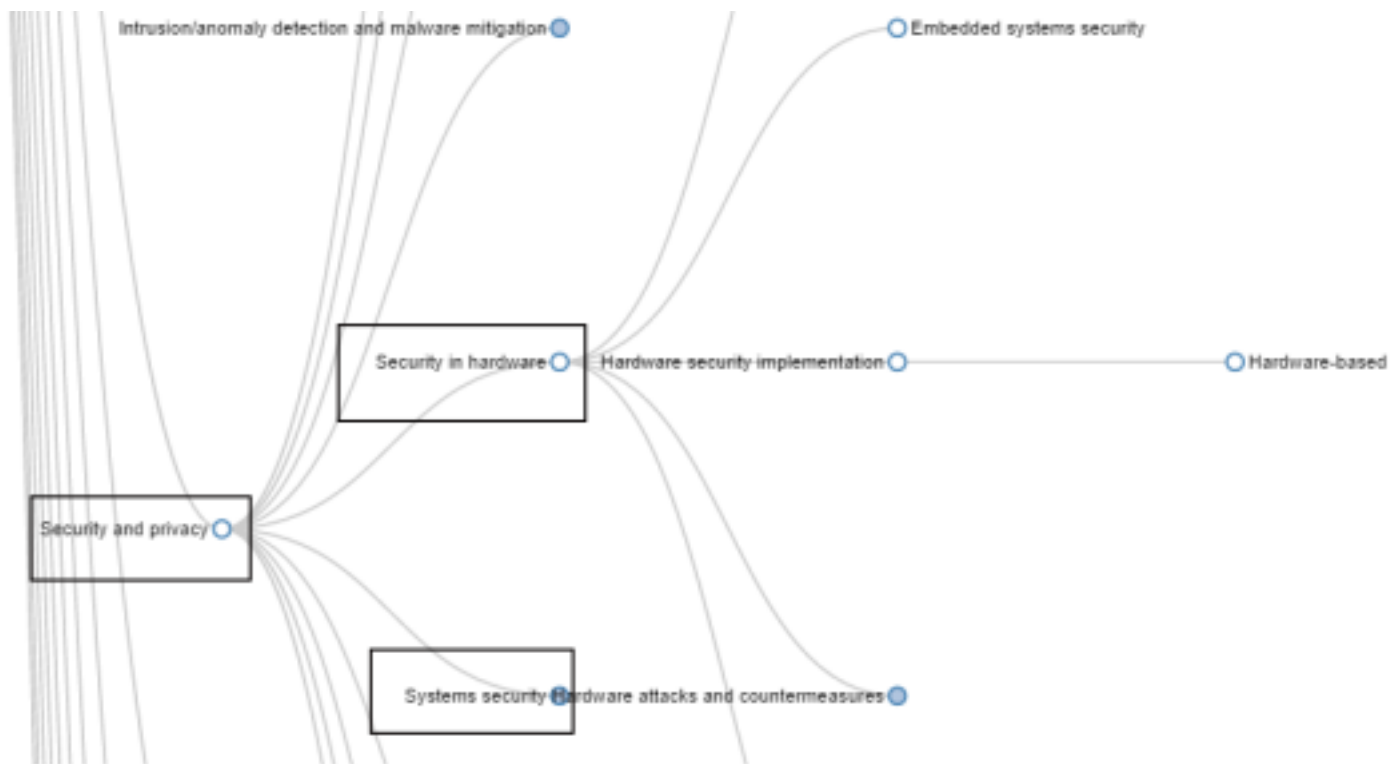


Figura 4.3: Trecho do Sistema de Classificação da ACM para áreas da Computação.

dos autores, caso não seja tratada na extração dos dados de entrada, podem causar a criação de vértices de autores adicionais gerando uma deformação no grafo da rede social. Os casos mais comuns dessa espécie de sinonímia aparecem como abreviações e supressões de parte do sobrenome dos pesquisadores, como é o caso da pesquisadora Alba Cristina Magalhães Alves de Melo que aparece como autora de publicações usando os nomes *Alba C. M. A. Melo* e *Alba de Melo*.

Para contornar a dificuldade na comparação de nomes, inspirando-se no trabalho de Digiampietri et al. (2015), porém usando um método menos custoso, uma vez que naquele trabalho são usadas muitas informações para realizar a desambiguação, optou-se por utilizar um método heurístico baseado em características extraídas dos nomes dos autores usando para isso a distância de edição, também conhecida por distância de *Levenshtein* e apresentada na Seção 2.4, conjugado com características da rede de colaboração entre os pesquisadores como vizinhos em comuns.

Tabela 4.1: Estatutos de periódicos conforme documento de área da CAPES (2013)

estrato	A1	A2	B1	B2	B3	B4	B5	C
$w$	100	85	70	50	20	10	5	0

Uma vez que a rede de colaboração científica tenha sido formada e que as áreas de pesquisa tenham sido identificadas, o próximo passo é identificar os centróides de cada área de pesquisa baseado na quantidade de publicações realizadas por cada pesquisador, bem como na recência dessas publicações, ou seja, quanto mais recente uma publicação mais ela contribuirá para o cálculo dos centróides, gerando dessa forma um fator de influência.

Tal influência fica exponencialmente menor quanto maior for a distância do seu ano de publicação (PY) com o ano base (BY), que normalmente se assume com o ano de consulta da base ou o ano no qual se está realizando a análise. O cálculo do fator de influência considera ainda o parâmetro  $w$  que representa o peso atribuído ao estrato do periódico na CAPES, conforme o documento de área de Ciência da Computação, exibido na Tabela 4.1 (CAPES, 2013). Para o cômputo do  $w$  foi desenvolvido um *script* em Python que recebe o nome do periódico e busca na tabela WebQualis<sup>4</sup> o estrato do periódico. Na ocorrência de abreviatura no nome do periódico, faz-se necessário uma nova consulta na versão *online* da base DBLP para obtenção do nome completo.

A hipótese para a utilização de um termo como o fator de influência é que, quanto mais recente uma publicação, mais relevante serão os conhecimentos utilizados pelo pesquisador para produzi-la, e pelo contrário, quanto mais antiga uma publicação maior será a probabilidade do pesquisador não estar mais atuando nesta área e, com isso, a área passa a ser menos relevante para ele. Algo similar acontece com o estrato que qualifica o periódico, quanto maior a classificação do periódico na CAPES mais trabalho deve ser empenhado na produção dessa publicação. O fator de influência, conforme calculado na Equação 4.1, é então usado para o cálculo de índice de recomendação (Rec) de cada pesquisador, conforme mostrado na Equação 4.2. Salienta-se que as recomendações geradas para cada pesquisador são realizadas por área de pesquisa identificada.

$$\rho_i = \left(\frac{1}{e^{BY-PY}}\right) * w \quad (4.1)$$

Considerando uma área específica de pesquisa, gera-se a recomendação para cada pesquisador calculando o *Rec* dele em relação aos demais. Desta forma é obtido um *ranking* dos pesquisadores. São excluídas as recomendações de parcerias já existentes, uma vez que objetiva-se a ampliação da rede de colaboração e não o reforço das parcerias existentes.

$$Rec = \sum_{i=1}^n \rho_i \quad (4.2)$$

## 4.2 Modelagem de dados

Na ausência de um modelo de dados NoSQL orientado a grafo amplamente aceito pela comunidade de banco de dados, foi utilizado um modelo de entidades, inspirado na topologia de grafos dirigidos, conforme apresentado na Figura 4.4. Note que os vértices representam entidades que compõem o modelo e as arestas representam os relacionamentos entre essas entidades. O modelo de dados em grafo foi adotado para melhor aproveitar as características do conceito de redes de colaboração que está sendo modelado cujo foco de estudo é a interação entre os vértices do tipo *Author*.

Os vértices *Author* estão relacionado a si mesmos em uma espécie de autorelacionamento, através da aresta *Recommended* que representa as recomendações de parcerias sugeridas, e ao vértice do tipo *Publication*, que representa as publicações realizadas, através

---

<sup>4</sup>Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf>

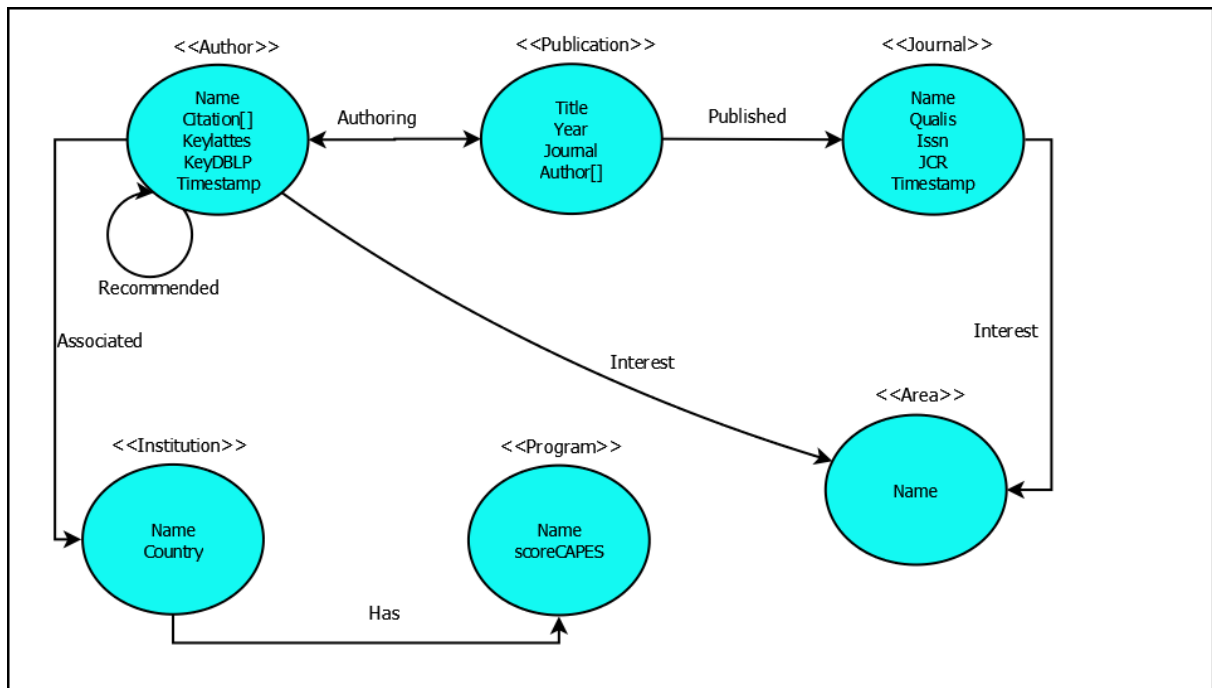


Figura 4.4: Modelo de dados usado para representar as entidades do modelo de recomendação

da aresta *Authoring*. O vértice *Author* está também relacionado com o vértice *Institution*, através da aresta *Associated*, que representa o relacionamento de vinculação entre um autor e sua universidade ou laboratório de pesquisa. O vértice *Publication* está ainda conectado ao vértice *Journal*, entidade usada para veiculação das publicações, e este com o vértice *Area* através da aresta *Interest*. O vértice *Area* está ainda conectado com o vértice *Author* para representar as áreas de pesquisa dos autores. Existe ainda a aresta *Has* que conecta os vértices *Institution* e *Program* e representa a vinculação de um programa de pós-graduação e uma universidade. Cabe destacar que a aresta *Authoring* é a única aresta bidirecional do grafo e essa característica permite que se possa partir de algum vértice *Author* e chegar em *Publication* e vice-versa.

### 4.3 Detalhamento da Implementação

O modelo de solução implementado será apresentado em dois níveis de abstração, conforme o *workflow* da Figura 4.1. A Figura 4.5 que descreve o processo de recuperação dos autores com suas respectivas produções, a carga do banco de dados NoSQL e a geração da rede social científica conforme passos 1 e 2 da Figura 4.1. A Figura 4.6 detalha o sub-processo de pré-processamento apresentado na Figura 4.5. A Figura 4.7 apresenta o processo de tratamento dos dados já carregados no grafo da rede de colaboração, para realizar a identificação dos agrupamentos de pesquisadores por área de pesquisa, geração das recomendações para cada autor e posterior cálculo das métricas de validação das recomendações geradas, conforme detalhado nos passos 3, 4 e 5 da Figura 4.1 do modelo de solução.

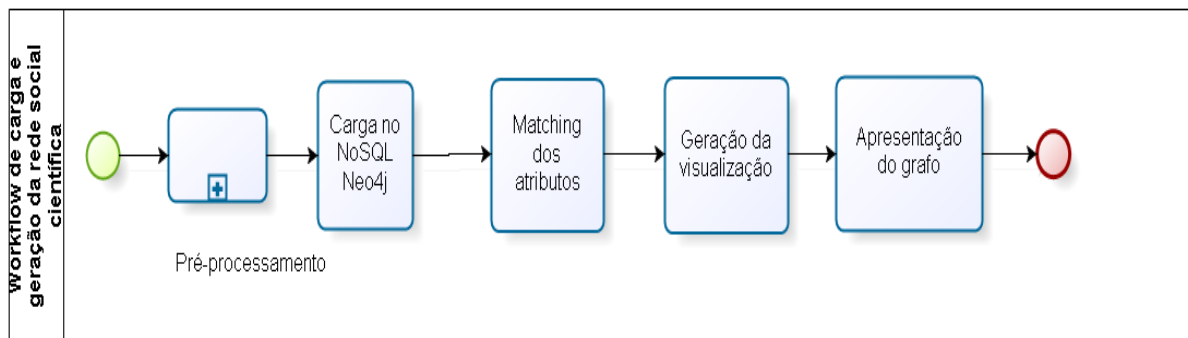


Figura 4.5: Processo de carga e visualização

A Figura 4.5 trata da carga dos dados e geração da rede de colaboração científica, incluindo desde a coleta dos nomes dos pesquisadores no repositório DBLP até a visualização da rede de colaboração científica como um grafo não dirigido. Note que o primeiro elemento é o sub-processo de pré-processamento (Figura 4.6), o qual inclui três atividades:

- **Obtenção da base DBLP:** O arquivo xml com toda a base DBLP é obtido a partir do próprio site da DBLP, no endereço <http://dblp.uni-trier.de/xml/>, compactado no formato gzip.
- **Obtenção da lista de pesquisadores:** Os nomes dos pesquisadores vinculados aos programas de pós-graduação selecionados são obtidos a partir das páginas web dos programas, sendo criado um arquivo JSON para cada programa na forma apresentada por (Mitchell, 2015). Para a realização dessa tarefa, foi criado um *script* em Python para cada programa de pós-graduação, o qual recebe o endereço da página com os nomes dos pesquisadores, e coleta além do nome, o endereço do currículo lattes e o email, quando disponível, criando o arquivo JSON. Os *scripts* criados estão disponíveis no Apêndice D.
- **Identificação das autorias:** São selecionadas as publicações que possuem na sua lista de autores algum autor vinculado aos programas de pós-graduação selecionados. Nessa atividade, os coautores também são identificados para posterior inserção no grafo da rede social. Aqui é realizada a normalização dos nomes dos autores, removendo ou substituindo caracteres fora da tabela ASCII e usando a distância de *Levenshtein*, definida na Seção 2.4, como heurística na comparação dos nomes dos autores.

Após o processo de pré-processamento, detalhado na Figura 4.6, a tarefa de carga no banco de dados em grafo Neo4j é iniciada, quando será realizado o *matching* dos atributos de autoria (nome para os vértices do tipo *Author* e autor para os vértices do tipo *Publication*) para criação da rede de colaboração científica. No total são carregados 3.534 vértices autores, incluindo os autores vinculados aos quatro programas de pós-graduação em Ciência da Computação (UFMG, USP, UFRN e UnB) e seus coautores, e 8.872 arestas. Na sequência, ocorre a geração da visualização, quando são construídos os gráficos da rede de colaboração. Posteriormente, o grafo da rede social científica está pronto para ser visualizado. O objetivo da tarefa de apresentação do grafo é permitir

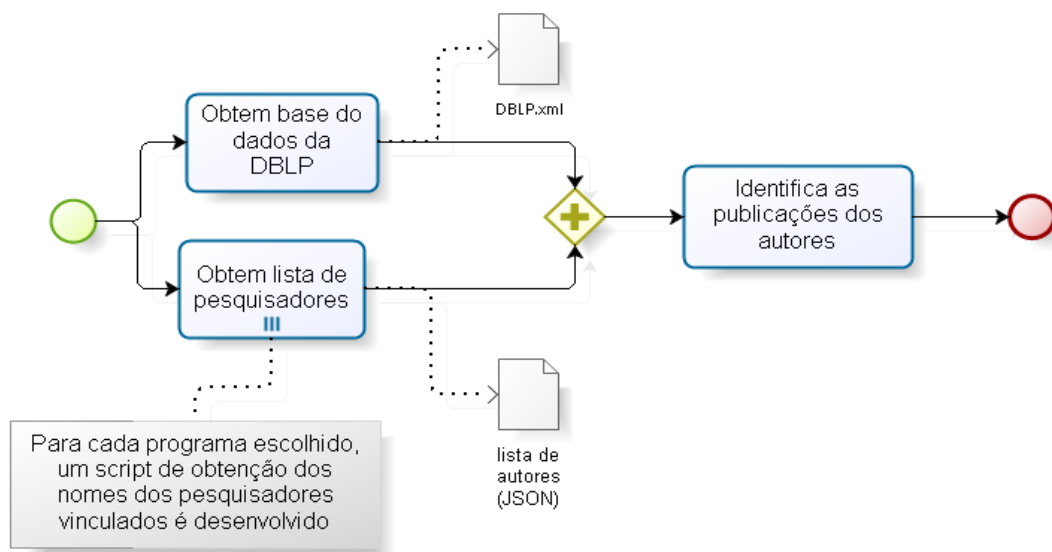


Figura 4.6: O processo de pré-processamento

a visualização dos relacionamentos entre os pesquisadores e o quão conectado um grupo de pesquisadores está com relação aos demais. Para auxiliar a visualização da rede a ferramenta de manipulação e visualização de grafos Gephi<sup>5</sup> foi adotada.

Uma vez que a rede social científica está carregada, é possível clusterizar os autores por área de pesquisa. O processo se inicia com a identificação dos títulos dos periódicos publicados por cada autor. Esses títulos são tratados com a retirada das palavras comuns (*stopwords*) e repetições entre os títulos do mesmo autor. A partir do título tratado, busca-se no sistema de classificação da ACM proximidade entre os literais aplicando a distância de *Levenshtein* (Equação 2.11). Forma-se então um grupo de áreas de pesquisa (*Bag of Areas*) por autor, conforme listagem disponível no Apêndice E.

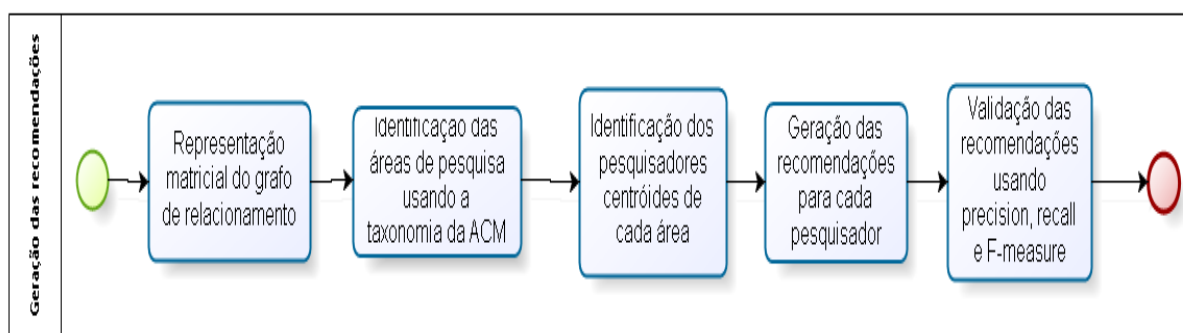


Figura 4.7: Processo de geração e validação das recomendações

A tarefa seguinte é a identificação dos centróides de cada área de pesquisa para, em seguida, gerar as recomendações de parcerias de pesquisa. Para o estudo de caso foi definido o parâmetro de cinco recomendações por pesquisador. No entanto, essa quantidade

<sup>5</sup>Disponível na url <http://www.gephi.org>

de recomendações pode ser alterada no *script* em Python de geração das recomendações, conforme listagem disponível no Apêndice F.

O processo encerra-se com a validação dos resultados usando as métricas de *Precision*, *Recall* e *F\_measure* adotando um período para as publicações e as publicações do ano seguinte ao período usadas como controle. Ressalta-se porém que esse intervalo de controle de um ano pode ser configurado no *script* do Apêndice F. Por exemplo, considere as publicações realizadas no período de 2005 até 2010. As publicações realizadas em 2011 serão utilizadas como controle para comprovar o desempenho das recomendações realizadas através do cálculo das métricas de *Precision*, *Recall* e *F\_measure*.

## 4.4 Tecnologias Utilizadas

A construção do artefato que implemente o modelo de recomendações demandou o uso de algumas tecnologias que auxiliaram a superar desafios da pesquisa. Um desses desafios foi o armazenamento da rede de colaboração em uma solução que aproveitasse os modelos de dados em grafo, para isso foi escolhido o banco de dados em grafo. Segundo comparativo realizado por Holzschuher and Peinl (2013) o Neo4j possui o melhor desempenho de consultas em modelos de dados baseados em grafos quando comparado com implementações que usam o modelo relacional.

Para a realização de cálculos analíticos e geração de grafos para teste, foi usada a biblioteca NetworkX Schult and Swart (2008), cuja implementação em Python Van Rossum et al. (2007) permitiu a integração com os artefatos de software desenvolvidos nessa pesquisa.

Também implementada em Python e com foco no processamento de linguagem natural, a biblioteca NLTK Bird (2006) foi usada na heurística para casamento de nomes de autores e de áreas de pesquisa. Para comunicação entre os vários módulos implementados, foi adotado o formato de dados Json (*Javascript Object Notation*) usado para representação de dados estruturado, sendo muito comum no desenvolvimento de serviços Web em detrimento do formato xml. Outra ferramenta utilizada foi o software de visualização e manipulação de grafos Gephi Bastian et al. (2009). Essas tecnologias serão detalhadas na sequência.

### Neo4j

Neo4j é um banco de dados NoSQL orientado a grafo multiflexível que pode operar com ou sem esquema, sob licença proprietária ou para comunidade, e em ambiente embarcado como servidor, permitindo ainda sua operação em um ambiente distribuído. Esse banco disponibiliza uma linguagem de consulta chamada Cypher cuja sintaxe é similar ao SQL.

Segundo Miller (2013), uma efetiva extração de informação de um grafo requer o que é conhecido como caminho no grafo, ou travessia. Um caminho no grafo consiste em visitar elementos do grafo. O caminho é uma operação fundamental na recuperação da informação. Uma grande diferença entre um caminho e uma consulta SQL é que caminhos são operações locais, ou seja, não existe índice de adjacência global de outra forma, no grafo, em que cada vértice e aresta armazena um mini-índice de objetos conectados a eles. Isso significa que o tamanho do grafo não impacta no desempenho do caminho e as custosas operações de agrupamento executadas através de JOINS na linguagem SQL se



tornam desnecessárias. É importante destacar que índices globais existem no Neo4j porém são usados apenas para agilizar a busca pelo ponto inicial de um caminho. Índices são requeridos para rapidamente recuperar vértices, baseado nas propriedades desses vértices. Eles proveem um ponto de partida para iniciar o caminho. Caso não existissem índices para determinar se um elemento particular tem uma certa propriedade, seria necessária uma busca linear por todos os elementos ao custo de  $O(n)$ , sendo  $n$  a quantidade de vértices no grafo. Alternativamente, o custo de uma busca em um índice é menor que  $O(\log_2 n)$ .

Um exemplo de uso de uma consulta em Cypher para recuperar os coautores e a quantidade de publicações com um autor é apresentado na Listagem 4.1. Essa consulta parte de um pesquisador cujo atributo *name* é “Nivio Ziviani” e a partir dos seus relacionamentos *AUTHORING* chega nos vértices do tipo *Publication*, e desse chega a outros vértices do tipo *Author*, que são os coautores em cada publicação. Uma vez que esses caminhos no grafo são identificados, são filtrados aqueles que possuem a publicação  $p$  com data de 2010 e são do tipo artigos em periódicos ou artigos em conferências. As informações retornadas da consulta são o nome dos coautores e a quantidade de publicações em comum desses autores, ordenadas de forma decrescente pela quantidade de publicações. O resultado da consulta na Listagem 4.1 pode ser visualizado na Tabela 4.2.

Listing 4.1: Retorna os coautores e a quantidade de publicações

**MATCH**

```
(a:Author {name:"Nivio_Ziviani"})-[r1:AUTHORING]-(p:Publication)-[r2:AUTHORING]-(b:Author)
```

**WHERE**

```
toInt(p.year) = 2010 AND p.type IN ['article', 'inproceedings']
```

**RETURN**

```
DISTINCT b.name AS coauthorName,
```

```
count (distinct p) as pubCount
```

```
ORDER BY pubCount DESC
```

Tabela 4.2: Resultado da consulta de coautores e a quantidade de publicações

coauthorName	pubCount
Edleno Silva de Moura	4
Marcos André Gonçalves	3
Wladmir C. Brandão	2
Thierson Couto	2
Marco Cristo	2
Jussara M. Almeida	2
Guilherme Vale Menezes	1
Berthier A. Ribeiro-Neto	1
Alberto H. F. Laender	1
Anísio Lacerda	1
Ricardo A. Baeza-Yates	1
Claudine Santos Badue	1

## NetworkX

De acordo com Schult and Swart (2008), a biblioteca NetworkX é uma ferramenta de análise de rede flexível escrita na linguagem Python. A NetworkX provê estruturas de dados baseadas em grafo, ou rede, que permitem a representação de grafos dirigidos contendo laços e arestas paralelas. Ela permite quase qualquer tipo de objeto como vértice e pode associar objetos arbitrários a arestas. Isto significa que a estrutura de rede pode ser composta por objetos customizados e estruturas de dados, complementando qualquer código pre-existente e permitindo análise de rede em qualquer aplicação sem muito esforço de desenvolvimento.

Uma vez que uma rede seja representada como um objeto NetworkX, algoritmos padrão para facilitar a busca pela distribuição de grau (número de arestas incidentes em cada vértice), coeficiente de agrupamento (número de triângulos que cada vértice é parte), caminhos mais curtos, medidas espectrais e comunidades podem ser usadas para analisar a estrutura do grafo. Dentre as opções disponibilizadas pela biblioteca NetworkX destacamos os algoritmos para cálculo do fluxo máximo entre dois vértices e a árvore geradora mínima do grafo da rede social científica.

Um exemplo de utilização da NetworkX para o cálculo do fluxo máximo entre dois vértices de um grafo é mostrado na Listagem 4.2, na qual  $G$  é o grafo extraído da rede de colaboração, `maximum_flow` é o método que calcula o fluxo máximo entre os vértices Pesquisador1 e Pesquisador2 e retorna o valor do fluxo máximo entre ambos (`flow_value`) e um novo grafo, na forma de um dicionário de listas de adjacências entre vértices, com as arestas de fluxo não-máximo removidas (`flow_dict`). Percorrendo todos os vértices do grafo, é possível calcular o fluxo máximo desse grafo como um todo. Uma particularidade desse método é a necessidade das arestas do grafo possuírem uma propriedade com nome `capacity` que armazena a capacidade de transmissão de fluxo dessa aresta.

Listing 4.2: Cálculo de fluxo máximo utilizando NetworkX

```
import networkx as nx
G = nx.Graph()
flow_value, flow_dict = nx.maximum_flow(G, Pesquisador1,
    Pesquisador2)
```

De forma semelhante, a criação da árvore geradora mínima de um grafo  $G$  seria conforme mostrado na Listagem 4.3, com o peso de cada aresta sendo representado pela propriedade `capacity`.

Listing 4.3: Criação da Árvore Geradora Mínima de um Grafo  $G$  pela NetworkX

```
import networkx as nx
G = nx.Graph()
MST = nx.minimum_spanning_tree(G, weight='capacity')
```

## NLTK

A biblioteca NLTK (*Natural Language Toolkit*) é escrita na linguagem Python, sendo muito útil no tratamento de textos. Nessa pesquisa, foi usado o método de cálculo de distancia de *Levenshtein*, conforme apresentado na Seção 2.4, para comparação heurística

entre nomes de pesquisadores, assim como o *corpus* para identificação de *stopwords* nos nomes dos periódicos usados para publicação pelos pesquisadores.

Bird (2006) descreve a NLTK como uma grande coleção de módulos minimamente interdependentes organizados em uma hierarquia com poucos níveis. Um conjunto de módulos centrais define tipos básicos, que são usados por toda a biblioteca, e os módulos restantes são “módulos de tarefas”, cada um com uma tarefa de processamento de linguagem natural específica. Por exemplo, o módulo *nlk.parser* é responsável por analisar a estrutura sintática de uma sentença, ou a *nlk.corpus* é responsável pelas tarefas referentes a *corpora* linguística.

Embora seja uma importante ferramenta para o tratamento de linguagem natural, por este não ser o foco deste trabalho, a ferramenta não foi explorada na sua plenitude.

## JSON

JSON (*JavaScript Object Notation*) serve para manipular dados usados em sistemas Web. O JSON foi usado originalmente para representar textualmente objetos da linguagem javascript. Nessa pesquisa, foi usado para armazenar a lista de nomes dos pesquisadores vinculados aos programas de pós-graduação considerados, como também para representar os dados usados pela interface Web da rede de colaboração. Um trecho de um dos arquivos de filtro de pesquisadores pode ser visualizado na Listagem 4.4 , note que trata-se de uma lista de objetos, cada um contendo os atributos *name*, *email* e *lattesurl*, que armazenam o nome, o email e o endereço do currículo Lattes, respectivamente, assim como disponibilizado na página Web do programa de pós-graduação.

Listing 4.4: Exemplo de filtro de pesquisadores

```
[
{"name": "Alba Cristina Magalhães Alves de Melo",
"email": "albamm@cic.unb.br",
"lattesurl": "http://lattes.cnpq.br/5123602572479168"},
{"name": "André C. Drummond ",
"email": "andred@cic.unb.br",
"lattesurl": "http://lattes.cnpq.br/8525132522308309"},
{"name": "Camilo C. Dorea",
"email": "camilo@cic.unb.br",
"lattesurl": "http://lattes.cnpq.br/7370796774582075"},
{"name": "Carla Denise Castanho ",
"email": "carlacastanho@cic.unb.br",
"lattesurl": "http://lattes.cnpq.br/2003154270448435"},
{"name": "Cláudia Nalon",
"email": "nalon@cic.unb.br",
"lattesurl": "http://lattes.cnpq.br/7793795625581127"},
{"name": "Célia Ghedini Ralha",
"email": "ghedini@cic.unb.br",
"lattesurl": "http://lattes.cnpq.br/5632722847264046"}
]
```

## Gephi

Gephi é uma ferramenta para manipulação e visualização de grafos. Com ela é possível realizar os principais cálculos em grafos e ainda visualizar numericamente e graficamente o resultado desses cálculos. Além disso, é possível aplicar filtros de vértices e arestas de forma a simplificar e até possibilitar a observação de determinadas características do grafo. O Gephi disponibiliza ainda uma interface para integração com bases de dados diversas, permitindo até mesmo que os dados armazenados no Neo4j sejam importados diretamente para ele. Algumas das métricas estatísticas que o Gephi disponibiliza são grau médio, grau médio ponderado, diâmetro da rede, densidade do grafo, HITS, modularidade, PageRank e a quantidade de Componentes conectados. Para o grafo gerado pela pesquisa e importado no Gephi, foram calculadas as métricas da Tabela 4.3. As métricas cujo valor estão marcadas com “\*” não geram um valor numérico único para todo o grafo, gerando valores para cada vértice, é o caso do HITS e do PageRank.

As métricas apresentadas são descritas da seguinte forma:

- Grau médio: Média dos graus de cada vértice.
- Grau médio ponderado: Média dos graus de cada vértice considerando o pesos de cada aresta.
- Diâmetro do grafo: Maior distância entre um par de vértices.
- Densidade do grafo: Relação entre quantidade de vértices e quantidade de arestas.
- HITS: Técnica que busca uma relação entre autoridades (vértices com conteúdo relevante sobre determinado tópico) e *hubs* (vértices que interligam as autoridades). O Gephi calcula a distribuição dos *hubs* e das autoridades no grafo.
- Modularidade: avalia o número de conexões intra-grupo em relação ao número esperado de arestas entre o mesmo grupo de vértices, porém em um grafo aleatório com a mesma sequência de graus (dos Santos and Rosset, 2015).
- PageRank: Segundo Page et al. (1999), é um método para calcular um *ranking* de páginas da Web. O Gephi calcula a distribuição do PageRank para cada vértice.
- Componentes conectados: Um componente conectado é um subgrafo maximal conexo de  $G$ , no qual cada vértice pertence a exatamente um componente conexo, e o mesmo é válido para as arestas.

O modelo de solução apresentado neste capítulo foi implementado em um artefato de *software* e, para verificar sua viabilidade, um estudo de caso foi conduzido. Os detalhes da implementação e a análise dos resultados obtidos com o estudo de caso serão apresentados no Capítulo 5.

Tabela 4.3: Cálculos estatísticos gerados pelo Gephi

Métrica	Valor
Grau médio	3,839
Grau médio ponderado	3,839
Diâmetro do grafo	11
Densidade do grafo	0,013
HITS	*
Modularidade	0,807
PageRank	*
Componentes conectados	20

# Capítulo 5

## Experimentação e Resultados

Visando validar o modelo desenvolvido foi realizado um estudo de caso usando como população da rede de colaboração os pesquisadores vinculados aos programas de pós-graduação da área de computação das instituições UFMG, USP, UFRN e UnB, com classificação na CAPES de 7, 6, 5 e 4 respectivamente. Na Seção 5.1 é descrito o estudo de caso conduzido; e na Seção 5.2 são discutidos os resultados envolvendo os aspectos quantitativos e qualitativos.

### 5.1 Estudo de Caso

O estudo de caso foi iniciado com a coleta dos nomes dos pesquisadores vinculados aos quatro programas de pós-graduação, sendo 47 da UFMG, 37 da USP, 29 da UFRN e 26 da UnB. O total inicial de 139 pesquisadores, adicionando-se os coautores desses, resultou em 3.158 autores inseridos na rede de colaboração. Esses autores produziram 3.055 publicações entre artigos em conferências (2.021), artigos em periódicos (1.021), livros (3) e coleções (10).

Paralelamente à coleta dos pesquisadores de cada programa de pós-graduação, foi obtida a base de publicações da DBLP, disponibilizada na forma de um único arquivo xml compactado no formato gzip de 323 MB de tamanho que, ao ser extraído, consome 1.8 GB, o que inviabilizava sua leitura por *parsers* de xml do tipo *Document Object Model* (DOM). O Apêndice B exibe um trecho do arquivo xml da DBLP.

Os dois processos de coleta dos dados e geração das recomendações estão apresentados na Figura 4.5 e em seguida as tarefas apresentadas na Figura 4.7 foram executadas para produzir as recomendações.

Foram produzidas 4.603 recomendações para 921 pesquisadores, com uma média de 5 e quantidade máxima de 8 recomendações por pesquisador. Buscou-se ainda garantir que todas as áreas de pesquisa automaticamente identificadas para um pesquisador contribuíssem com pelo menos uma recomendação.

Para exemplificar a geração das recomendações, tomemos a pesquisadora Célia Ghedini Ralha cujas áreas de pesquisa identificadas foram as seguintes:

- Decision support systems;
- Distributed artificial intelligence;

- Spatial-temporal systems;
- Computational biology.

Das áreas de pesquisa identificadas, foram geradas as recomendações constantes na Tabela 5.1. Dessalta-se que as recomendações não possuem relação de ordem entre elas. Em cada recomendação produzida, constam o nome do pesquisador recomendado, a qual instituição esse pesquisador está vinculado. Ressalta-se que essa informação está presente apenas para aqueles pesquisadores vinculados aos programas de pós-graduação foco deste estudo. Constam também a bolsa de produtividade em pesquisa do CNPq, se o pesquisador tiver, além da área de pesquisa em cuja recomendação se originou. O questionário de avaliação das recomendações respondido pela pesquisadora é o primeiro apresentado no Apêndice G. Note que referente as recomendações da Tabela 5.1 foram aceitos três recomendações entre as cinco primeiras da tabela (página 77). A geração dessas recomendações pode ser reproduzida acessando o endereço <http://scicoll-academicmetrics.rhcloud.com/recommending>.

Tabela 5.1: Exemplo de recomendações de parceria

Nome	Instituição	Bolsa PQ	Área de pesquisa
Maria Emilia Machado Telles Walter	UnB	PQ-2	Computational biology
Ronaldo Fumio Hashimoto	USP	PQ-2	Computational biology
Wagner Meira Jr.	UFMG	PQ-1C	Spatial-temporal systems
Antonio Alfredo Ferreira Loureiro	UFMG	PQ-1A	Spatial-temporal systems
Maurício Ayala Rincón	UnB	PQ-1D	Distributed artificial intelligence
Jaime Simão Sichman	USP	PQ-2	Distributed artificial intelligence
Alba Cristina Magalhães Alves de Melo	UnB	PQ-1D	Decision support systems
Arnaldo de Albuquerque Araújo	UFMG	PQ-1D	Decision support systems

## 5.2 Análise dos Resultados

O estudo de caso conduzido foi avaliado quantitativamente e qualitativamente. A avaliação quantitativa foi realizada por meio do cálculo das métricas de *precision*, *recall* e *F\_measure* das recomendações produzidas, com os resultados sendo exibidos na Tabela 5.2. Foram consideradas as publicações nos anos de 2010 a 2014 como dados e 2015 como controle, portanto não apresentam dados nas três primeiras colunas da última linha da tabela.

A partir dos dados presentes na Tabela 5.2, observa-se que no ano de 2012 o modelo de recomendação apresentou o melhor desempenho. Parte desse desempenho pode ser atribuído ao grande incremento na quantidade de publicações e nas relações de coautoria ocorridos no ano de 2013 (ano usado como controle para as recomendações de 2012). A evolução anual das métricas pode ser melhor visualizada na Figura 5.1. Em uma tentativa

de explicar o salto na quantidade de publicações ocorrido nos anos de 2013 e 2014, pode-se atribuir à avaliação dos cursos de pós-graduação de Computação realizada pela CAPES no ano de 2013.

Tabela 5.2: Resultado das métricas de avaliação das recomendações

Ano	Precision	Recall	F_measure	Periódicos	Coautoria
2010	0.010206	0.328671	0.019798	63	284
2011	0.012378	0.335294	0.023874	73	299
2012	0.054723	0.980545	0.103661	77	318
2013	0.029099	0.629108	0.055625	104	454
2014	0.022584	0.630303	0.043606	108	331
2015				80	331

Ressalta-se contudo que embora os dados de publicações já realizadas em 2016 estejam presentes na base de dados, esses não foram considerados por não apresentarem o quantitativo de um ano completo.

A avaliação qualitativa foi realizada através da aplicação via Web de um questionário para um subgrupo de pesquisadores que compõem a rede social. O questionário aplicado está disponível no Apêndice A. O critério de seleção dos pesquisadores foi a vinculação a qualquer uma das quatro instituições usadas para criação da rede de colaboração. Foi utilizado o grupo de pesquisadores para os quais foi possível identificar automaticamente as áreas de pesquisa, conforme os veículos de publicação.

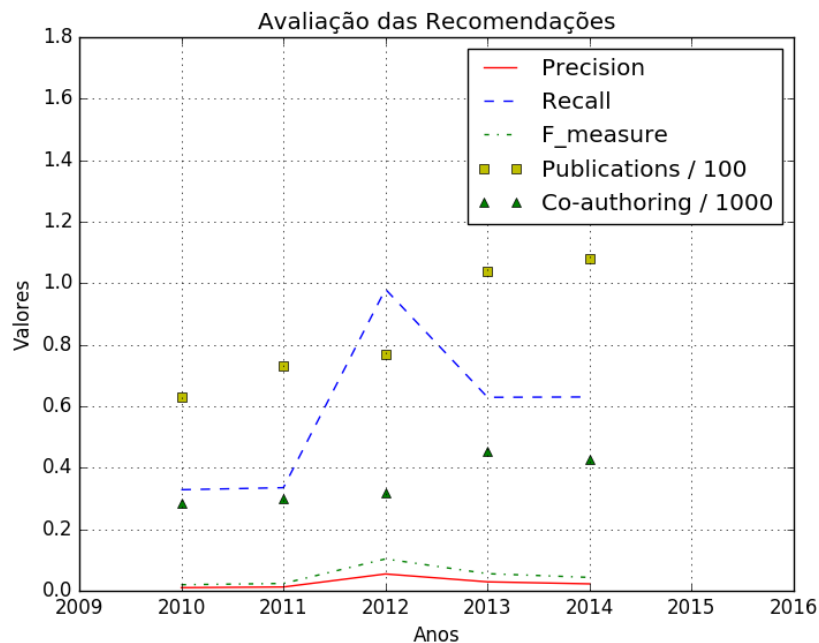


Figura 5.1: Evolução anual das métricas de avaliação das recomendações

No total foram enviados 71 questionários, tendo sido respondidos 16. A coleta das respostas ocorreu no período de 29/05/2016 à 15/06/2016. Os questionários respondidos podem ser visualizados no Apêndice G.



A Tabela 5.3 apresenta um resumo das análises das respostas do questionário. Note que o Sistema de Classificação de áreas da Computação da ACM, apresentou um índice de aceitação de 58%, porém ao se solicitar uma sugestão de sistema de classificação as respostas foram: “Sim”, “deveria ter IA”.

Tabela 5.3: Síntese das respostas ao questionário de avaliação

Critério	Valor
Aceita o Sistema de Classificação da ACM	10
Não aceita o Sistema de Classificação da ACM	2
Aceita Parcialmente o Sistema de Classificação da ACM	4
Aceita a DBLP como repositório de publicações	4
Número de participantes	17
Quantidade de recomendações geradas	79
Quantidade de recomendações aceitas	25
Precisão das recomendações	31%

Pode-se concluir que, embora esteja longe de ser uma unanimidade, não foi possível identificar uma alternativa ao Sistema de Classificação da ACM para subáreas da Computação.

Quando solicitados a deixar um comentário sobre a base de publicações usada nessa pesquisa as respostas informadas indicam uma postura neutra quanto ao uso da DBLP, com as seguintes respostas:

- Desconheço a DBLP.
- Eu questiono um pouco o DBLP, porque não tem tanto controle de qualidade.
- Apesar da base DBLP não ser completa, ela é curada.
- DBLP, como o nome diz, é focado em subáreas específicas, e é deficiente em outras. Usar como base uma base de dados enviesada favorecerá algumas áreas e desfavorecerá outras de forma permanente.

Na análise das áreas de pesquisa declaradas pelos pesquisadores, um quarto delas estão em português, mesmo tendo sido solicitado para usar o Sistema de Classificação da ACM. Esse fato inviabiliza o casamento automático com as áreas do Sistema de Classificação da ACM que usa o inglês como único idioma. A Figura 5.2 exibe um mapa de palavras das áreas de pesquisa indicadas pelos pesquisadores, na qual é possível perceber que as áreas de maior destaque são *software* e *system*. Note que a existência de dois idiomas (português e inglês) para descrever as áreas de pesquisa dificulta o tratamento automático e pode gerar ruído nas análises.

As recomendações geradas e avaliadas pelos pesquisadores totalizaram 79, ou seja, esse é o tamanho do subconjunto das recomendações geradas cujos pesquisadores responderam ao questionário. Das recomendações geradas, 25 foram aceitas e 54 foram rejeitadas,



# Capítulo 6

## Conclusões e Trabalhos Futuros

Com a conclusão dessa etapa da pesquisa, foi possível responder a questão de pesquisa acerca da possibilidade de se definir e implementar um artefato para análise de redes de colaboração científica que realize recomendações de parcerias meritorias para pesquisadores. O modelo desenvolvido para geração de recomendações de parcerias e sua implementação na forma de um sistema de informação demonstraram a viabilidade de responder tal questão, comprovada pelas avaliações quantitativas e qualitativas realizadas através da condução do estudo de caso. O percentual de aceitação das recomendações geradas foi de 31% conforme apresentado na Seção 5.2. Apesar do baixo índice de respondentes do questionário (16 entre 71), considera-se um bom resultado, conforme literatura da área de sistemas de recomendação (média abaixo de 30%), segundo Debnath et al. (2008).

O modelo de recomendação de parcerias desenvolvido pode assumir um papel importante na busca por novas parcerias em pesquisas científicas, permitindo a realização dessas parcerias usando como elemento chave as publicações já realizadas pelos pesquisadores agrupadas por áreas de pesquisa. A sugestão é viabilizada por meio da recomendação de parcerias empregada como ferramenta de melhoria nos índices de publicações científicas, mostrando-se 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.

Sabe-se porém que, para que uma parceria em pesquisa científica seja efetivamente realizada, um conjunto de fatores sociais são extremamente relevantes, embora não tenha sido possível a identificação desse conjunto através de um método analítico. Para tanto, se faz necessária uma análise mais profunda da dimensão social dos participantes da rede de colaboração científica através de um modelo mais complexo.

Como trabalhos futuros pretende-se melhorar o desempenho das recomendações geradas, aumentando a índice de aceitação das recomendações produzidas, assim como expandir a base de dados utilizada de forma a abranger todos os pesquisadores da área de Computação no Brasil que possuam publicações na DBLP. Otimizar o desempenho da geração das recomendações, tendo em vista hoje não ser possível realizar um conjunto de recomendações imediatamente após um pesquisador ser inserido na rede de colaboração, é outra possibilidade de trabalhos futuros. Pretende-se ainda experimentar outras abordagens no modelo de recomendação aumentando a quantidade de camadas de relacionamentos do modelo da rede de colaboração, no qual cada camada representa uma forma de interação entre pesquisadores como coautoria em artigos ou periódicos, participação

em projeto apenas para citar algumas possibilidades.

# Referências

- Adolfo Alonso-Arroyo, Antonio Pulgarín, and Isidoro Gil-Leiva. Estudio cuantitativo de la colaboración científica en la universidad politécnica de valencia (españa). *Information Research*, 11(1), 2005. 3
- Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Recuperação de Informação - 2ed: Conceitos e Tecnologia das Máquinas de Busca*. Bookman Editora, second edition, 2013. ISBN 9788582600498. 12, 13, 14
- Renato Balancieri, Alessandro Botelho Bovo, Vinícius Medina Kern, Roberto Carlos dos Santos Pacheco, and Ricardo Miranda Barcia. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na plataforma lattes. *Revista IBICT*, 34(1), 2005. 2
- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999. 2, 16
- Albert-Laszlo Barabási, Hawoong Jeong, Zoltan Neda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3):590–614, 2002. 2, 15
- Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009. URL <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>. 36
- Jorge Biolchini, Paula Gomes Mian, Ana Candida Cruz Natali, and Guilherme Horta Travassos. Systematic review in software engineering. Technical Report 679/05, COPPE/UFRJ, May 2005. 22
- Steven Bird. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics, 2006. 36, 39
- CAPES. Documento de Área 2013 - ciência da computação. <http://www.capes.gov.br/component/content/article?id=4656:ciencia-da-computacao>, Oct 2013. xii, 31, 32
- Aurélio Ribeiro Costa and Célia Ghedini Ralha. Análise de rede de colaboração científica como ferramenta na gestão de programas de pós-graduação. *SBSI*, 70:970, 2015a. 11

- Aurélio Ribeiro Costa and Célia Ghedini Ralha. Análise de rede de colaboração científica como ferramenta na gestão de programas de pós-graduação. In *II Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI-SBSI2015)*, pages 13–15, May 2015b. 11
- Blaise Cronin. *The hand of science: Academic writing and its rewards*. Scarecrow Press, 2005. 3
- Allison Davis, Burleigh Bradford Gardner, and Mary R Gardner. *Deep South: A social anthropological study of caste and class*. Univ of South Carolina Press, 2009. 14
- Gerald F. Davis and Henrich R. Greve. Corporate elite networks and governance changes in the 1980s. *American journal of sociology*, 103(1):1–37, 1997. 14
- Souvik Debnath, Niloy Ganguly, and Pabitra Mitra. Feature weighting in content based recommendation system using social network analysis. In *Proceedings of the 17th international conference on World Wide Web*, pages 1041–1042. ACM, 2008. 4, 47
- Luciano Digiampietri, Ricardo Linden, and Lenin Barbosa. Desambiguação de nomes em redes sociais acadêmicas: Um estudo de caso usando dblp. In *IV Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2015)*, 2015. 27, 31
- Edsger Wybe Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 12 1959. doi: 10.1007/BF01386390. 25
- Camila Pereira dos Santos and Mariá Cristina Vasconcelos Nascimento Rosset. Agrupamento em grafos direcionados pela maximização da modularidade ajustada. *Simpósio Brasileiro de Pesquisa Operacional*, 2015. 40
- Roberto Carlos dos Santos Pacheco and Vinícius Medina Kern. Uma ontologia comum para a integração de bases de informações e conhecimento sobre ciência e tecnologia. *Ci. Inf*, 30(3):56–63, 2001. 28
- Michael D Ekstrand, John T Riedl, and Joseph A Konstan. Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, 4(2): 81–173, 2011. 18, 19
- Lester R Ford and Delbert R Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8(1):399–404, 1956. doi: 10.4153/CJM-1956-045-5. 27
- Vinícius P. Freire and Daniel R. Figueiredo. Ranking in collaboration networks using a group based metric. *Journal of Brazillian Computer Society*, 17(4):255–266, November 2011. 15
- Joseph Galaskiewicz and Peter V Marsden. Interorganizational resource networks: Formal patterns of overlap. *Social science research*, 7(2):89–107, 1978. 14
- Kailash C. Garg and Pitambar Padhi. A study of collaboration in laser science and technology. *Scientometrics*, 51(2):415–427, 2001. 3

- Wolfgang Glänzel. National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1):69–115, 2001. 3
- Wolfgang Glänzel. Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. *Library trends*, 50(3):461–475, 2002. 3
- David Goldberg, David Nichols, Brian M Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992. 19
- Waleej Haider, Nouman M Durrani, Shardha Nand, Nadeem Kafi Khan, and M Asad Abasi. Towards knowledge creation and management model over online social networks. In *Web Applications and Networking (WSWAN), 2015 2nd World Symposium on*, pages 1–6. IEEE, 2015. 12
- Jon Herlocker, Joseph A Konstan, and John Riedl. An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. *Information retrieval*, 5(4):287–310, 2002. 20
- Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237. ACM, 1999. 20
- Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in information systems research. *MIS Q.*, 28(1):75–105, March 2004. ISSN 0276-7783. URL <http://dl.acm.org/citation.cfm?id=2017212.2017217>. 7, 11
- Florian Holzschuher and René Peinl. Performance of graph query languages: comparison of cypher, gremlin and native access in neo4j. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, pages 195–204. ACM, 2013. 36
- J Sylvan Katz and Ben R Martin. What is research collaboration? *Research policy*, 26(1):1–18, 1997. 2, 3
- Leo Katz. A new status index derived from sociometric index. *Psychometrika*, pages 39–43, 1953. 27
- Ki-Wan Kim. Measuring international research collaboration of peripheral countries: Taking the context into consideration. *Scientometrics*, 66(2):231–240, 2006. 3
- Barbara Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26, 2004. 22
- Manfred Kochen. *The small world*. Ablex Pub., 1989. 16
- Chen Luo, Wei Pang, Zhe Wang, and Chenghua Lin. Hete-cf: Social-based collaborative filtering recommendation using heterogeneous relations. In *2014 IEEE International Conference on Data Mining*, pages 917–922. IEEE, 2014. 27

- Paula Mählck and Olle Persson. Socio-bibliometric mapping of intra-departmental networks. *Scientometrics*, 49(1):81–91, 2000. 3
- Maria de Fátima S. Maia, Pesquisas Epidemiológicas da UFPel, and Sônia Elisa Caregnato. Co-autoria como indicador de redes de colaboração científica. *Perspectivas em Ciência da Informação*, 13(2):18–31, 2008. 2, 3
- PETER MARIOLIS. Interlocking directorates and control of corporations: The theory of bank control. *Social Science Quarterly*, 56(3):425–439, 1975. ISSN 00384941, 15406237. URL <http://www.jstor.org/stable/42859564>. 14
- Christopher Marshall, Pearl Brereton, and Barbara Kitchenham. Tools to support systematic reviews in software engineering: A feature analysis. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, EASE '14*, pages 13:1–13:10, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2476-2. doi: 10.1145/2601248.2601270. URL <http://doi.acm.org/10.1145/2601248.2601270>. 22
- Andrew McAfee, Erik Brynjolfsson, Thomas H Davenport, DJ Patil, and Dominic Barton. Big data. *The management revolution. Harvard Bus Rev*, 90(10):61–67, 2012. 12
- Arthur Jack Meadows and Antonio Agenor Briquet de Lemos. *A comunicação científica*. Briquet de Lemos/livros, 1999. 3
- Prem Melville and Vikas Sindhvani. *Recommender systems*, pages 829–838. Springer, 2010. doi: [http://dx.doi.org/10.1007/978-0-387-30164-8\\_705](http://dx.doi.org/10.1007/978-0-387-30164-8_705). 18, 19
- Martin Meyer and Sujit Bhattacharya. Commonalities and differences between scholarly and technical collaboration. *Scientometrics*, 61(3):443–456, 2004. 3
- Justin J Miller. Graph database applications and concepts with neo4j. In *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA*, volume 2324, 2013. 36
- Ryan Mitchell. *Web scraping with Python: collecting data from the modern web*. "O'Reilly Media, Inc.", 2015. 34
- Gonzalo Navarro. A guided tour to approximate string matching. *ACM computing surveys (CSUR)*, 33(1):31–88, 2001. 21
- Mark EJ Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical review E*, 64(1):016131, 2001. 14
- Mark EJ Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences*, 101(suppl 1):5200–5205, 2004. 3
- Wanda J Orlikowski and C Suzanne Iacono. Research commentary: Desperately seeking the "it" in it research—a call to theorizing the it artifact. *Information systems research*, 12(2):121–134, 2001. 7
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999. URL <http://ilpubs.stanford.edu:8090/422/>. 40



- Paul Resnick and Hal R Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997. 18
- Daniel A Schult and P Swart. Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*, volume 2008, pages 11–16, 2008. 36, 38
- Upendra Shardanand and Pattie Maes. Social information filtering: algorithms for automating "word of mouth". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217. ACM Press/Addison-Wesley Publishing Co., 1995. 20
- Hiroshi Shimodaira. Similarity and recommender systems. *School of Informatics, The University of Eidenburgh*, 2014. 21
- 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. 46
- Herbert A Simon. *The sciences of the artificial*, volume 136. MIT press, 1996. xi, 6, 7, 8, 10
- Victor Ströele, Ricardo Silva, Moisés Ferreira de Souza, Carlos Eduardo R. de Mello, Jano M. Souza, and Geraldo Zimbrão. Identifying workgroups in brazilian scientific social networks. *Journal of Universal Computer Science*, 17(14):1951–1970, Oct 2011a. 25
- Victor Ströele, Geraldo Zimbrão, and Jano M. Souza. Modeling, mining and analysis of multi-relational scientific social network. *Journal of Universal Computer Science*, 18(8):1048–1068, Apr 2011b. 27
- 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. ISSN 0164-1212. doi: 10.1016/j.jss.2013.02.024. 15, 19, 27
- Maria Inês Tomaél and Regina Maria Marteleto. Redes sociais: posições dos atores no fluxo da informação 10.5007/1518-2924.2006 v11nesp1p75. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 11(1):75–91, 2007. 4
- Hiroyuki Tomizawa and Masashi Shirabe. Likelihood of overseas access to international co-authorships. *Scientometrics*, 53(1):113–121, 2002. 3
- 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. doi: 10.1201/9781420059335. URL <http://www.crcnetbase.com/doi/abs/10.1201/9781420059335>. 11
- Guido Van Rossum et al. Python programming language. In *USENIX Annual Technical Conference*, volume 41, 2007. 36

- Samile Andrea de Souza Vanz and Ida Regina Chittó Stumpf. Colaboração científica: revisão teórico conceitual. *Perspectivas em Ciência da Informação*, 15(2):42–55, 2010. 2
- Stanley Wasserman and Kathrine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994. 16, 27
- Ron Weber. Toward a theory of artifacts: A paradigmatic base for information systems research. *Journal of Information Systems*, 1(2):3–19, 1987. 8
- Ron Weber. Still desperately seeking the it artifact. *MIS quarterly*, 27(2):183–183, 2003. 8
- Bo Yang, Yu Lei, Dayou Liu, and Jiming Liu. Social collaborative filtering by trust. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2747–2753. AAAI Press, 2013. 27
- Fuyuki Yoshikane and Kyo Kageura. Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, 60(3):435–446, 2004. 3
- Xiao Yu, Xiang Ren, Quanquan Gu, Yizhou Sun, and Jiawei Han. Collaborative filtering with entity similarity regularization in heterogeneous information networks. *IJCAI HINA*, 2013. 27
- Zhenxin Yu, Hong Yan, and TC Edwin Cheng. Benefits of information sharing with supply chain partnerships. *Industrial management & Data systems*, 101(3):114–121, 2001. 4
- Augusto B. Zamboni, André D. Thommazo, Elis Cristina M. Hernandez, and Sandra Carmargo P. F. Fabbri. Start uma ferramenta computacional de apoio à revisão sistemática. In *Proc.: Congresso Brasileiro de Software (CBSOft'10), Salvador, Brazil*, 2010. 22

# Apêndice A

## Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável neste link (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ? \_\_\_\_\_

Sugeriria outro sistema de classificação ? \_\_\_\_\_

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

1 - \_\_\_\_\_

2 - \_\_\_\_\_

3 - \_\_\_\_\_

4 - \_\_\_\_\_

5 - \_\_\_\_\_

6 - \_\_\_\_\_

7 - \_\_\_\_\_

8 - \_\_\_\_\_

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados DBLP como fonte de referência na área de Computação ? \_\_\_\_\_

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

\_\_\_\_\_  
\_\_\_\_\_

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ? \_\_\_\_\_

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- Área de pesquisa \_\_\_\_\_
- Quantidade de periódicos qualificados \_\_\_\_\_
- Vínculo institucional \_\_\_\_\_
- Possibilidade de interação presencial \_\_\_\_\_
- Se existe a possibilidade de projeto financiado \_\_\_\_\_
- Bom relacionamento pessoal \_\_\_\_\_
- Pesquisador PQ/CNPq \_\_\_\_\_
- Possibilidade de oferta de bolsas \_\_\_\_\_
- Outro, qual ? \_\_\_\_\_

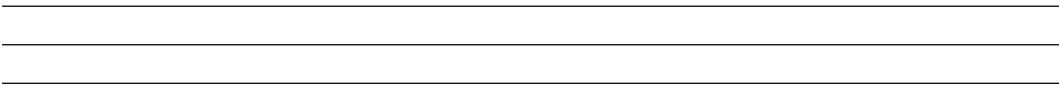
6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 - \_\_\_\_\_ [área de pesquisa: \_\_\_\_\_]  
2 - \_\_\_\_\_ [área de pesquisa: \_\_\_\_\_]  
3 - \_\_\_\_\_ [área de pesquisa: \_\_\_\_\_]  
4 - \_\_\_\_\_ [área de pesquisa: \_\_\_\_\_]  
5 - \_\_\_\_\_ [área de pesquisa: \_\_\_\_\_]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

- Recomendação 1 - Área: Sim \_\_\_\_\_ Não \_\_\_\_\_. Comentário: \_\_\_\_\_
- Recomendação 2 - Área: Sim \_\_\_\_\_ Não \_\_\_\_\_. Comentário: \_\_\_\_\_
- Recomendação 3 - Área: Sim \_\_\_\_\_ Não \_\_\_\_\_. Comentário: \_\_\_\_\_
- Recomendação 4 - Área: Sim \_\_\_\_\_ Não \_\_\_\_\_. Comentário: \_\_\_\_\_
- Recomendação 5 - Área: Sim \_\_\_\_\_ Não \_\_\_\_\_. Comentário: \_\_\_\_\_

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.



Obrigado pela colaboração.

# Apêndice B

## Trecho da Base de Dados DBLP

Este é um trecho do arquivo xml disponibilizado pela DBLP<sup>1</sup>, o arquivo original contém todos autores e publicações disponibilizadas online.

```
<article mdate="2003-11-25" key="journals/acta/Csuhaj-VarjuM00">
<author>Erzs&eacute;bet Csuhaj-Varj&uacute;</author>
<author>Victor Mitrana</author>
<title>Evolutionary Systems: A Language Generating Device Inspired by
  Evolving Communities of Cells.</title>
<pages>913-926</pages>
<year>2000</year>
<volume>36</volume>
<journal>Acta Inf.</journal>
<number>11</number>
<ee>
http://link.springer.de/link/service/journals/00236/bibs/0036011/00360913.htm
</ee>
<url>db/journals/acta/acta36.html#Csuhaj-VarjuM00</url>
</article>
```

```
<article mdate="2008-05-21" key="journals/acta/Janicki08">
<author>Ryszard Janicki</author>
<title>Relational structures model of concurrency.</title>
<pages>279-320</pages>
<year>2008</year>
<volume>45</volume>
<journal>Acta Inf.</journal>
<number>4</number>
<ee>http://dx.doi.org/10.1007/s00236-008-0071-6</ee>
<url>db/journals/acta/acta45.html#Janicki08</url>
</article>
```

```
<article mdate="2011-01-11" key="journals/acta/Russell77">
```

---

<sup>1</sup>Endereço do xml <http://dblp.uni-trier.de/xml/>

<author>Bruce Russell</author>  
<title>On an Equivalence between Continuation and Stack Semantics.</title>  
<journal>Acta Inf.</journal>  
<volume>8</volume>  
<year>1977</year>  
<pages>113-123</pages>  
<url>db/journals/acta/acta8.html#Russell177</url>  
<ee><http://dx.doi.org/10.1007/BF00289244></ee>  
</article>

<article mdate="2011-01-11" key="journals/acta/EhrenfeuchtR74">  
<author>Andrzej Ehrenfeucht</author>  
<author>Grzegorz Rozenberg</author>  
<title>Nonterminals Versus Homomorphisms in Defining Languages for Some  
Classes of Rewriting Systems.</title>  
<journal>Acta Inf.</journal>  
<volume>3</volume>  
<year>1974</year>  
<pages>265-283</pages>  
<url>db/journals/acta/acta3.html#EhrenfeuchtR74</url>  
<ee><http://dx.doi.org/10.1007/BF00288638></ee>  
</article>

<article mdate="2011-01-11" key="journals/acta/Kemp80">  
<author>Rainer Kemp</author>  
<title>A Note on the Density of Inherently Ambiguous Context-free Languages.</title>  
<pages>295-298</pages>  
<year>1980</year>  
<volume>14</volume>  
<journal>Acta Inf.</journal>  
<url>db/journals/acta/acta14.html#Kemp80</url>  
<ee><http://dx.doi.org/10.1007/BF00264258></ee>  
</article>

<article mdate="2011-01-11" key="journals/acta/HanI89">  
<author>Yijie Han</author>  
<author>Yoshihide Igarashi</author>  
<title>Time Lower Bounds for Parallel Sorting on a Mesh-Connected  
Processor Array.</title>  
<pages>643-655</pages>  
<year>1989</year>  
<volume>26</volume>  
<journal>Acta Inf.</journal>  
<number>7</number>  
<url>db/journals/acta/acta26.html#HanI89</url>  
<ee><http://dx.doi.org/10.1007/BF00288975></ee>

</article>



# Apêndice C

## Áreas do Sistema de Classificação da ACM

Lista com as áreas de pesquisa utilizadas pelos pesquisadores da rede de colaboração. Essas áreas foram automaticamente identificadas a partir do título dos periódicos utilizados.

Applied computing

- Life and medical sciences
- Bioinformatics
- Computational biology
- Enterprise computing
- Enterprise information systems

Information systems

- Data management systems
- Database management system engines
- Database administration
- Information systems applications
- Spatial-temporal systems
- Decision support systems
- Information retrieval
- Retrieval models and ranking
- World Wide Web
- Web applications
- Information storage systems
- Storage management

Software and its engineering

- Software organization and properties
- Software system structures
- Software notations and tools
- General programming languages
- Software creation and management
- Designing software

## Networks

- Network types
  - Ad hoc networks
  - Overlay and other logical network structures
- Network algorithms
  - Control path algorithms
- Network properties
  - Network structure

## Computing methodologies

- Artificial intelligence
  - Computer vision
  - Distributed artificial intelligence
  - Knowledge representation and reasoning
- Symbolic and algebraic manipulation
  - Symbolic and algebraic algorithms
- Modeling and simulation
  - Simulation types and techniques

## Human-centered computing

- Collaborative and social computing
  - Collaborative and social computing theory, concepts and paradigms
  - Collaborative and social computing systems and tools
- Human computer interaction (HCI)
  - Interaction paradigms
  - Interaction devices

## Hardware

- Communication hardware, interfaces and storage
  - Signal processing systems
- Robustness
  - Hardware reliability

## Social and professional topics

- Professional topics
  - Management of computing and information systems

## Social and professional topics

- Professional topics
  - Computing education
  - History of computing
  - Computing and business

## Mathematics of computing

- Probability and statistics
  - Probabilistic representations

- Statistical paradigms
- Probabilistic reasoning algorithms
- Probabilistic inference problems
- Discrete mathematics
- Graph theory

Computer systems organization

- Architectures
- Distributed architectures
- Other architectures

Security and privacy

- Systems security
- Operating systems security

Theory of computation

- Theory and algorithms for application domains
- Algorithmic game theory and mechanism design
- Models of computation
- Concurrency

# Apêndice D

## Scripts de Coleta dos Pesquisadores

Listing D.1: Script para coleta dos dados dos pesquisadores da UFMG

```
from bs4 import BeautifulSoup as bs
from urllib.request import urlopen

tiposProf = ['CO', 'PE', 'PA']
docentes=list()

for tpProf in tiposProf:
    soup =bs(urlopen('https://www.dcc.ufmg.br/pos/pessoas/
        professores2.php?tipo=%s' % tpProf).read())
    profTbl=soup.find_all(attrs={'valign':'top', 'class':'
        td_sub_conteudo'})

    indice=1
    for profCell in profTbl:
        try:
            if tpProf == 'PE':
                docentes.append('"name":"' + profCell.strong.
                    text + '"')
                indice +=1
        except:
            pass

f = open("docentes-ufmg.json", "w")
f.write('[')
for i in docentes:
    f.write("{%s}," % i)

f.write('null|')
f.close()
```

Listing D.2: Script para coleta dos dados dos pesquisadores da USP

```
from bs4 import BeautifulSoup as bs
```

```

from urllib.request import urlopen

url = 'http://www.ime.usp.br/dcc/pos/orientadores'
uspResearchers = list()

soup = bs(urlopen(url))
maincolumn = soup.find('div', attrs={'id': 'maincolumn'})

for tr in maincolumn.find_all('tr'):
    td = tr.find_all('td')

    if len(td) == 0:
        continue
    td_lattes = td[3].a['href']

    uspResearchers.append('"name":_' + td[0].text + '_, "lattes_url":_'
        + td_lattes + '"')

f = open('docentes-usp.json', 'w')
f.write('{ "usp": [')
for i in uspResearchers:
    f.write("{%s},\n" % i)
f.write('null] }')
f.close()

```

Listing D.3: Script para coleta dos dados dos pesquisadores da UFRN

```

#!/*- coding: utf-8 -*-
from bs4 import BeautifulSoup as bs
from urllib.request import urlopen
import sys

baseurl = u"https://sigaa.ufrn.br/sigaa/public/programa/equipe.jsf?lc=pt_BR&id=73"
professors=list()
soup =bs(urlopen(baseurl).read(), from_encoding='utf-8')

table = soup.find(id="table_lt")
if not table:
    print("No_table_found")
    sys.exit(-1)
for tr in table.find_all('tr'):
    if tr['class'] == 'campos':
        continue
    row = tr.find_all('td')
    a_tag = row[4].find('a')
    if a_tag:

```

```

        professors.append({"name":row[0].text.strip(), "lattes":
            a_tag['href']})
    else:
        professors.append({"name":row[0].text.strip(), "lattes":
            ""})

with open("docentes-ufrn.json", 'w') as f:
    f.write('[')
    for p in professors:
        if p['name'] == 'Nome':
            continue
        f.write('{"name":"%s","lattesurl":"%s"},' % (p['name'],p
            ['lattes']))
    f.write('null]')

```

Listing D.4: Script para coleta dos dados dos pesquisadores da UnB

```

#!/usr/bin/python
from bs4 import BeautifulSoup
from urllib.request import urlopen
import sys, re

print("Iniciando parser da página de docentes")

class Researcher:
    def __init__(self, name='', email='', lattesurl=''):
        self.name = name
        self.email = email
        self.lattesurl = lattesurl

    def __repr__(self):
        return '{"name":_"' + self.name + '",_ "email":_' +
            ' + self.email + '",_ "lattesurl":_' + self.
            lattesurl + '"}\n'

try:
    pagina = urlopen("http://ppgi.unb.br/curso/docentes")
except Exception as e:
    print("Erro ao ler a página_%s", str(e))
    sys.exit(1)

soup = BeautifulSoup(pagina)
docentes = list()
tabela_docentes = soup.find_all(attrs={"class": "docentes"})

contador = 1

```

```

for docente in tabela_docentes[0].find_all(attrs={"style": "
background-color: #d8e8e9;"}):
    links = docente.find_all("a")
    nome = links[0]
    lattes = links[1]['href']
    email = links[2].text.replace("_[at]_", "@")

    docentes.append(Researcher(nome.text, email, lattes))
    contador += 1
print(docentes)

f = open("docentes.json", 'w')
f.write(docentes.__str__())
f.close()

```

# Apêndice E

## Script Bag of Areas

Listing E.1: Script para criação do grupo de áreas de pesquisa

```
from py2neo import Graph, Node
from nltk.corpus import stopwords
import pickle, json

from DBLP.ACMAreas2 import Concept, ConceptRepo

cachedStopWords = stopwords.words('english')
customStopWords = ['Journal', 'IEEE', 'ACM', '(JASIST)']

def processBag(pseudobag):
    """
    @description Receive a list of journals title, extract
    stopwords and black listed
    words and merge them in a list.
    @param pseudobag list of journals where author has published
    @return a list words extracted from journals title
    """
    bag = set()

    for journal in pseudobag:
        wordsList = '_'.join(
            [word for word
             in journal.replace('[', '_').replace(']', '_').
             split()
             if word.lower() not in cachedStopWords
             and word not in customStopWords]
        )
        bag.add(wordsList)

    return bag

def loadACMareas(inJson = False):
```



```

'''
@description Load a python pickled file from a preprocessed
  ACM classification
system xml.
'''
acmfile = 'DBLP\\ACMconcepts.p'

repoAreas = pickle.load(open(acmfile, 'rb'))
if inJson is True:
    jsonAreas = repoAreas.jsonify()
    return json.loads(jsonAreas)
else:
    return repoAreas

def simpleMatch(str1, str2):
'''
Match the two strings if at least two words correnpond
'''
countmatch = 0
for i in str1.split():
    if str2.lower().find(i.lower()) >=0:
        countmatch += 1

return countmatch

def matchAreasTree(str1, tree):
l3_roots = [subSubArea for area in tree['children'] for
    subArea in area['children'] for subSubArea in subArea['
    children']]

matchedRootName = ''
for area in l3_roots:
    if area.get('children') is None:
        if area['name'] == str1:
            matchedRootName = area['name']
            break
    else:
        if simpleMatch(str1, str(area['children'])) >= 2:
            matchedRootName = area['name']
            break

return matchedRootName

#####
print("Starting_process..._be_patient")

```

```

g = Graph("http://datagraph-academicmetrics.rhcloud.com:80/db/
data")

aName = list()
bags = dict()
print("Creating_authors_list")
aCount = 0
for a in g.find("Author"):
    # Skip authors with bags, this allows incremental running
    if a['bagofareas'] is not None:
        continue

    pubs = [x.end_node for x
             in a.match_outgoing(rel_type="AUTHORING")
             if x.end_node['type'] == 'article']
    if len(pubs) == 0:
        #print("Author %s has no article published" % a['name'])
        continue
    aCount += 1
    print("(%i)_Trying_to_create_bag_of_areas_for_" % aCount,
          a['name'],
          "_for_",
          str(len(pubs)),
          "_articles")

    pubsl = sorted(pubs, key=lambda p: p['year'], reverse=True)

    pseudobag = set()
    for pub in pubsl:
        pubTitle = pub['title']
        journal, = [x.end_node for x in pub.match_outgoing(
            rel_type="Published")]

        pseudobag.add(journal['title'].split(',')[0])
        if len(pseudobag) > 3:
            break

    bags[a] = processBag(pseudobag)

print("Loading_ACM_areas")
jsonRepo = loadACMareas(True)

print("Report")
count = 0
for author in bags.keys():
    matchedareas = list()

```

```

for area in bags[author]:

    ma = matchAreasTree(area , jsonRepo)
    if ma != '':
        if ma not in matchedareas:
            matchedareas.append(ma)
if len(matchedareas) > 0:
    count += 1
    print ("%i)_%s_atua_nas_areas_ACM:_%s" %(count ,
                                                author[ 'name' ] ,
                                                matchedareas)

    )
    print ("_*_*_%s\n"%bags[author])
    author[ 'bagofareas' ] = matchedareas
    author.push()

print ("Finished")

```

# Apêndice F

## Script de Geração das Recomendações

Listing F.1: Script para geração das recomendações

```
from py2neo import Graph, Relationship
import operator
import time, datetime, math

minYear = 1990
year = 2010
MAXREC = 5.0
pubTypes = [ 'article' ]

startTime = time.time()

def haveContributed(g, author1, author2):
    retVal = g.cypher.execute(
        '''MATCH (a:Author {name:"%s"})-[r1]-(p:Publication)-[r2]
            (b:Author {name:"%s"})
WHERE toInt(p.year) <= %i AND p.type IN %s
RETURN count(distinct p) as pubs'''
        %(author1['name'], author2, year, pubTypes)
    )

    if retVal.one > 0:
        return True
    else:
        return False

def precision(recConfirmed, recDone):
    '''
    @param recConfirmed: Recommendations that were actually done
    @param recDone: Number of given recommendations
    '''
    return recConfirmed / recDone
```

```

def recall(recConfirmed , publications):
    '''
    @param recConfirmed: Recommendations that were actually done
    @param publications: Total of publications actually done
    '''
    if publications == 0.0:
        return 0.0
    return recConfirmed / publications

def f_measure(precision , recall):
    if precision == 0.0 and recall == 0.0:
        return 0.0
    f_measure = 2 *((precision * recall ) / (precision + recall)
    )

    return f_measure

def clearRecomendations(graph):
    print("Cleaning_previous_recommendations", end='_ ', flush=
        True)
    for rel in graph.match(rel_type="RECOMMENDATION"):
        graph.delete(rel)
    print("done_!")

#####
# Remoto
#graph = Graph("http://datagraph-academicmetrics.rhcloud.com:80/
    db/data")

#Local
graph = Graph()

authorList = list()
rank_of_areas = dict()

print("Looking_for_authors_to_generate_ranking")
for author in graph.find("Author"):
    authoring = 0
    bagofareas = author['bagofareas']
    if bagofareas is None:
        continue

    for j in author.match(rel_type="AUTHORING"):
        authoring += 1
    author['authoring'] = authoring
    if j.end_node['type'] in pubTypes:

```

```

    authorList.append(author)
for area in bagofareas:
    if rank_of_areas.get(area) is None:
        rank_of_areas[area] = [({'name': author['name'], '
            pubCount': authoring})]
    else:
        rank_of_areas[area].append({'name': author['name'],
            'pubCount': authoring})

print("Sorting_ranking")

for area in rank_of_areas.keys():
    rank_of_areas[area] = sorted(rank_of_areas[area],
        key = lambda x:x['pubCount'],
        reverse = True
    )

print("Recomendation_section")

authorRec = {}
clearRecomendations(graph)

for nodeAuthor in graph.find('Author'):
    name = nodeAuthor['name']

    if nodeAuthor['bagofareas'] is None:
        continue
    areas = nodeAuthor['bagofareas']
    print("Generating_recommendations_for:_", name)

    recomendations = []
    maxRec_normalized = math.ceil(MAXREC / len(areas))
    for area in areas:
        authors_in_area = rank_of_areas[area]
        recCount = 0
        for author in authors_in_area:
            if recCount >= maxRec_normalized:
                break

            if author['name'] == name:
                continue
            if haveContributed(graph, nodeAuthor, author['name']
                ]):
                continue

        recomendations.append(author['name'])

```

```

        rec_node = graph.find_one("Author", 'name', author[ '
            name' ])
        rel = Relationship(nodeAuthor, "RECOMMENDATION",
            rec_node, at=str(time.time()))
        rel.properties['area'] = area
        graph.create(rel)

        recCount +=1
        authorRec[name]=recomendations

print("\\nAll_recomendations_have_been_loaded_on_authorRec\\n")

# Time to validate de recomendations
totalResearchers = 0
countRec = 0
countPubs = 0
# Clear recomendations from graph

for author in authorRec.keys():
    totalResearchers += 1
    friends = graph.cypher.execute(
        '''MATCH (a:Author {name:"%s"})-[r1]-(p:Publication)-[r2
            ]-(b:Author)
WHERE toInt(p.year) = %i AND p.type IN %s
RETURN DISTINCT b.name AS friendName, count (distinct p) as
pubCount'''
        %(author, year+1, pubTypes)
    )
    publicationCount = (friends[0].pubCount if len(friends) > 0
        else 0)

    countPubs += publicationCount

confirmedRec = {}
for p in authorRec[author]:
    for f in friends:
        if f.friendName == p:
            if not f.friendName in confirmedRec.keys():
                confirmedRec[f.friendName] = 1
                countRec +=1
            else:
                confirmedRec[f.friendName] += 1
if len(confirmedRec) > 0:
    print("Recoms_confirmed_from_%s_(%i)_to_%s" %(author,
        publicationCount, confirmedRec))
print('\\n')

```

```

print("Total_of_researchers_(authors_of_articles):_",
      totalResearchers)

print("Publicacoes:_%i" %countPubs)
precision = precision(countRec, totalResearchers * MAXREC)
recall = recall(countRec, countPubs)
f_measure = f_measure(precision, recall)
print("Number_of_confirmed_recomendations:_%i_(%f_%%)" %(
      countRec, countRec/totalResearchers * 100) )

print("=====")
print("Precision:_%f" % precision)
print("Recall:_%f" % recall)
print("F-measure:_%f" % f_measure)
print("=====")
elapsedTime = time.time() - startTime
print("Tempo_decorrido:_%s_" %(datetime.timedelta(seconds=
      elapsedTime)))

```



Apêndice G

Questionários de Avaliação

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                                     |
|-----|-------------------------------------|
| 1 - | Information Systems                 |
| 2 - | Decision Support Systems            |
| 3 - | Artificial Intelligence             |
| 4 - | Distributed Artificial Intelligence |
| 5 - | Multi-Agent System                  |
| 6 - | Multi-Agent System                  |
| 7 - | Multi-Agent System                  |
| 8 - | Multi-Agent System                  |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="9"/>  | Quantidade de periódicos qualificados |
| <input type="text"/>            | Vínculo institucional                 |
| <input type="text"/>            | Possibilidade de interação presencial |

- 6  Se existe a possibilidade de projeto financiado  
 7  Bom relacionamento pessoal  
 8  Pesquisador PQ/CNPq  
 5  Possibilidade de oferta de bolsas  
 Outro, qual ?

Clique aqui para adicionar critério

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]  
 2 -  [área de pesquisa:  
 ]  
 3 -  [área de pesquisa:  
 ]  
 4 -  [área de pesquisa:  
 ]  
 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Maria Emilia Telles Walter - <b>UNB</b> - Area: Computational biology

Sim  Não. Comentário:

Ronaldo Fumio Hashimoto (PQ-2) - <b>USP</b> - Area: Computational biology

Sim  Não. Comentário:

Wagner Meira Jr. - Area: Spatial-temporal systems

Sim  Não. Comentário:

Antonio Alfredo Ferreira Loureiro (PQ-1A) - <b>UFMG</b> - Area: Spatial-temporal systems

Sim  Não. Comentário:

Maurício Ayala Rincón (PQ-1D) - <b>UNB</b> - Area: Distributed artificial intelligence

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |  |
|-----|--|
| 1 - | Collaborative and social computing systems |
| 2 - | Collaborative and social computing         |
| 3 - | Open Source Software                       |
| 4 - | Software and its engineering               |
| 5 - |  |
| 6 - |  |
| 7 - |  |
| 8 - |  |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="7"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="4"/>  | Vínculo institucional                 |
| <input type="text" value="6"/>  | Possibilidade de interação presencial |

- 9  Se existe a possibilidade de projeto financiado
  - 8  Bom relacionamento pessoal
  - 5  Pesquisador PQ/CNPq
  - 3  Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]
- 2 -  [área de pesquisa:  
 ]
- 3 -  [área de pesquisa:  
 ]
- 4 -  [área de pesquisa:  
 ]
- 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Célia Ghedini Ralha (PQ-2) - **UNB** - Area: Spatial-temporal systems

Sim  Não. Comentário:

Sérgio Vale Aguiar Campos (PQ-2) - **UFMG** - Area: Spatial-temporal systems

Sim  Não. Comentário:

Azzedine Boukerche - Area: Spatial-temporal systems

Sim  Não. Comentário:

Wagner Meira Jr. - Area: Spatial-temporal systems

Sim  Não. Comentário:

Antonio Alfredo Ferreira Loureiro (PQ-1A) - **UFMG** - Area: Spatial-temporal systems

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                           |
|-----|---------------------------|
| 1 - | proof theory              |
| 2 - | automated reasoning       |
| 3 - | modal and temporal logics |
| 4 - | constructive mathematics  |
| 5 - | description logics        |
| 6 - | description logics        |
| 7 - | description logics        |
| 8 - | description logics        |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="0"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="5"/>  | Vínculo institucional                 |
| <input type="text" value="5"/>  | Possibilidade de interação presencial |

- 0 Se existe a possibilidade de projeto financiado  
 10 Bom relacionamento pessoal  
 0 Pesquisador PQ/CNPq  
 0 Possibilidade de oferta de bolsas  
Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]  
2 -  [área de pesquisa:  
 ]  
3 -  [área de pesquisa:  
 ]  
4 -  [área de pesquisa:  
 ]  
5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Flavia Coimbra Delicato - <b>UFRN</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Mario Fernando Montenegro Campos (PQ-1B) - <b>UFMG</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Uira Kulesza (PQ-2) - <b>UFRN</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Maria Emília Telles Walter - <b>UNB</b> - Area: Computational biology

Sim  Não. Comentário:

Célia Ghedini Ralha (PQ-2) - <b>UNB</b> - Area: Computational biology

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |  |
|-----|--|
| 1 - | Multi-agent systems                    |
| 2 - | Knowledge representation and reasoning |
| 3 - | Interactive learning environments      |
| 4 - | Learning management systems            |
| 5 - | Accessibility                          |
| 6 - | Accessibility                          |
| 7 - | Accessibility                          |
| 8 - | Accessibility                          |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="5"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="7"/>  | Vínculo institucional                 |
| <input type="text" value="7"/>  | Possibilidade de interação presencial |



- 9  Se existe a possibilidade de projeto financiado
- 10  Bom relacionamento pessoal
- 7  Pesquisador PQ/CNPq
- 8  Possibilidade de oferta de bolsas
- Outro, qual ?

Clique aqui para adicionar critério

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]
- 2 -  [área de pesquisa:  
 ]
- 3 -  [área de pesquisa:  
 ]
- 4 -  [área de pesquisa:  
 ]
- 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Mario Paolucci - Area: Algorithmic game theory and mechanism design

- Sim  Não. Comentário:

Rosaria Conte - Area: Algorithmic game theory and mechanism design

- Sim  Não. Comentário:

Benjamin Rene Callejas Bedregal (PQ-1D) - **UFRN** - Area: Distributed artificial intelligence

- Sim  Não. Comentário:

Maurício Ayala Rincón (PQ-1D) - **UNB** - Area: Distributed artificial intelligence

- Sim  Não. Comentário:

Ana Paula Couto da Silva (PQ-2) - **UFMG** - Area: Software system structures

- Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                                    |
|-----|------------------------------------|
| 1 - | Concurrent computing methodologies |
| 2 - | computer systems organisation      |
| 3 - | Embedded hardware                  |
| 4 - |                                    |
| 5 - |                                    |
| 6 - |                                    |
| 7 - |                                    |
| 8 - |                                    |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="3"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="8"/>  | Vínculo institucional                 |
| <input type="text" value="8"/>  | Possibilidade de interação presencial |

- 7  Se existe a possibilidade de projeto financiado
  - 10  Bom relacionamento pessoal
  - 1  Pesquisador PQ/CNPq
  - 7  Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]
- 2 -  [área de pesquisa:  
 ]
- 3 -  [área de pesquisa:  
 ]
- 4 -  [área de pesquisa:  
 ]
- 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Mirella Moura Moro (PQ-2) - <b>UFMG</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Raquel Oliveira Prates - <b>UFMG</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Jaime Simão Sichman (PQ-2) - <b>USP</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Mario Fernando Montenegro Campos (PQ-1B) - <b>UFMG</b> - Area: Database management system engines

Sim  Não. Comentário:

Nívio Ziviani (PQ-1A) - <b>UFMG</b> - Area: Database management system engines

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                                   |
|-----|-----------------------------------|
| 1 - | Inteligência Artificial           |
| 2 - | Lógica Formal                     |
| 3 - | Pesquisa Operacional              |
| 4 - | Ontologias                        |
| 5 - | Sistemas baseados em Conhecimento |
| 6 - | Sistemas baseados em Conhecimento |
| 7 - | Sistemas baseados em Conhecimento |
| 8 - | Sistemas baseados em Conhecimento |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                |                                       |
|--------------------------------|---------------------------------------|
| <input type="text" value="9"/> | Área de pesquisa                      |
| <input type="text" value="9"/> | Quantidade de periódicos qualificados |
| <input type="text" value="8"/> | Vínculo institucional                 |
| <input type="text" value="8"/> | Possibilidade de interação presencial |

- 10 Se existe a possibilidade de projeto financiado  
 9 Bom relacionamento pessoal  
 Pesquisador PQ/CNPq  
 Possibilidade de oferta de bolsas  
Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]  
2 -  [área de pesquisa:  
 ]  
3 -  [área de pesquisa:  
 ]  
4 -  [área de pesquisa:  
 ]  
5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Fernando Magno Quintão Pereira (PQ-2) - <b>UFMG</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Jacir Luiz Bordim - <b>UNB</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Flavia Coimbra Delicato - <b>UFRN</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Mario Fernando Montenegro Campos (PQ-1B) - <b>UFMG</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Uira Kulesza (PQ-2) - <b>UFRN</b> - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                                   |
|-----|-----------------------------------|
| 1 - | database management system engine |
| 2 - | XML query languages               |
| 3 - | stream management                 |
| 4 - | Online social networks            |
| 5 - |                                   |
| 6 - |                                   |
| 7 - |                                   |
| 8 - |                                   |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |    |                                       |
|----|---------------------------------------|
| 10 | Área de pesquisa                      |
| 1  | Quantidade de periódicos qualificados |
| 1  | Vínculo institucional                 |
| 3  | Possibilidade de interação presencial |

- 1  Se existe a possibilidade de projeto financiado
  - 8  Bom relacionamento pessoal
  - 1  Pesquisador PQ/CNPq
  - 1  Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]
- 2 -  [área de pesquisa:  
 ]
- 3 -  [área de pesquisa:  
 ]
- 4 -  [área de pesquisa:  
 ]
- 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Lyrene Fernandes Da Silva - <b>UFRN</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Mariza Andrade da Silva Bigonha - <b>UFMG</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Marcio Eduardo Kreutz - <b>UFRN</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Raquel Oliveira Prates - <b>UFMG</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Jaime Simão Sichman (PQ-2) - <b>USP</b> - Area: Probabilistic representations

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.



# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

1 -	Parallel Algorithms
2 -	Distributed Algorithms
3 -	Bioinformatics
4 -	
5 -	
6 -	
7 -	
8 -	

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

8	Área de pesquisa
9	Quantidade de periódicos qualificados
2	Vínculo institucional
7	Possibilidade de interação presencial

- 7 Se existe a possibilidade de projeto financiado
  - 7 Bom relacionamento pessoal
  - 7 Pesquisador PQ/CNPq
  - 6 Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  ]
- 2 -  [área de pesquisa:  ]
- 3 -  [área de pesquisa:  ]
- 4 -  [área de pesquisa:  ]
- 5 -  [área de pesquisa:  ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Ana Paula Couto da Silva (PQ-2) - **UFMG** - Area: Bioinformatics

Sim  Não. Comentário:

Wagner Meira Jr. - Area: Bioinformatics

Sim  Não. Comentário:

Nívio Ziviani (PQ-1A) - **UFMG** - Area: Database management system engines

Sim  Não. Comentário:

Marcos André Gonçalves (PQ-1D) - **UFMG** - Area: Database management system engines

Sim  Não. Comentário:

Marco Cesar Goldbarg (PQ-2) - **UFRN** - Area: Decision support systems

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

- Sim  
 Não  
 Parcialmente. Por que ? IA é uma tópic importante

Sugeriria outro sistema de classificação ? deveria ter IA

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                         |
|-----|-------------------------|
| 1 - | Theory of computation   |
| 2 - | Network and Web         |
| 3 - | Information systems     |
| 4 - | Security and privacy    |
| 5 - | Artificial intelligence |
| 6 - | Artificial intelligence |
| 7 - | Artificial intelligence |
| 8 - | Artificial intelligence |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

9

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |   |                                       |
|---|---------------------------------------|
| 9 | Área de pesquisa                      |
| 9 | Quantidade de periódicos qualificados |
| 9 | Vínculo institucional                 |
| 8 | Possibilidade de interação presencial |

- 9  Se existe a possibilidade de projeto financiado
  - 9  Bom relacionamento pessoal
  - 10  Pesquisador PQ/CNPq
  - 8  Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 - Ana Bazzan  [área de pesquisa:  
 ]
- 2 - André Ponce de Leon F. de Carvalho  [área de pesquisa:  
 ]
- 3 - Dimitris Bertsimas  [área de pesquisa:  
 ]
- 4 - Fábio Gagliardi Cozman  [área de pesquisa:  
 ]
- 5 - James Schummer  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Jussara M. Almeida - Area: Management of computing and information systems

Sim  Não. Comentário:

Nívio Ziviani (PQ-1A) - <b>UFMG</b> - Area: Management of computing and information systems

Sim  Não. Comentário:

Marcos André Gonçalves (PQ-1D) - <b>UFMG</b> - Area: Management of computing and information systems

Sim  Não. Comentário:

Sim  Não. Comentário:

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                         |
|-----|-------------------------|
| 1 - | Systems biology         |
| 2 - | Bioinformatics          |
| 3 - | Biological networks     |
| 4 - | Machine learning theory |
| 5 - | Regression analysis     |
| 6 - | Regression analysis     |
| 7 - | Regression analysis     |
| 8 - | Regression analysis     |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |    |                                       |
|----|---------------------------------------|
| 10 | Área de pesquisa                      |
| 9  | Quantidade de periódicos qualificados |
| 7  | Vínculo institucional                 |
| 5  | Possibilidade de interação presencial |

- 5  Se existe a possibilidade de projeto financiado  
9  Bom relacionamento pessoal  
5  Pesquisador PQ/CNPq  
5  Possibilidade de oferta de bolsas  
Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  Robert Kaas [área de pesquisa:  
 Probability and statistics ]  
2 -  Robert Tibshirani [área de pesquisa:  
 Machine learning theory ]  
3 -  Marina Meila [área de pesquisa:  
 Unsupervised learning and clustering ]  
4 -  Trevor Hastie [área de pesquisa:  
 Machine learning theory ]  
5 -  Bradley Efron [área de pesquisa:  
 Machine learning theory ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Ana Paula Couto da Silva (PQ-2) - <b>UFMG</b> - Area: Bioinformatics

Sim  Não. Comentário:

Wagner Meira Jr. - Area: Bioinformatics

Sim  Não. Comentário:

Alba Cristina Magalhães Alves de Melo (PQ-1D) - <b>UNB</b> - Area: Bioinformatics

Sim  Não. Comentário:

Mohammed Javeed Zaki - Area: Statistical paradigms

Sim  Não. Comentário:

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

- Sim  
 Não  
 Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |   |
|-----|---|
| 1 - | Parallel Computing Methodologies        |
| 2 - | Models of Computation                   |
| 3 - | Computational Geometry                  |
| 4 - | Computer Systems Organization           |
| 5 - | Development Frameworks and Environments |
| 6 - | Development Frameworks and Environments |
| 7 - | Development Frameworks and Environments |
| 8 - | Development Frameworks and Environments |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="6"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="8"/>  | Vínculo institucional                 |
| <input type="text" value="8"/>  | Possibilidade de interação presencial |

- 6  Se existe a possibilidade de projeto financiado
  - 8  Bom relacionamento pessoal
  - 4  Pesquisador PQ/CNPq
  - 4  Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]
- 2 -  [área de pesquisa:  
 ]
- 3 -  [área de pesquisa:  
 ]
- 4 -  [área de pesquisa:  
 ]
- 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

- Joel H. Saltz - Area: General programming languages  
 Sim  Não. Comentário:
- Tahsin M. Kurç - Area: General programming languages  
 Sim  Não. Comentário:
- Tulio Tavares - Area: General programming languages  
 Sim  Não. Comentário:
- George Teodoro - Area: General programming languages  
 Sim  Não. Comentário:
- Renato Ferreira - Area: General programming languages  
 Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.



# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

- Sim  
 Não  
 Parcialmente. Por que ? Algumas áreas não são cor

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |  |
|-----|--|
| 1 - | Verificação formal                     |
| 2 - | Bioinformática                         |
| 3 - | Vídeo sob Demanda                      |
| 4 - | Sistemas de gestão de dados biológicos |
| 5 - |  |
| 6 - |  |
| 7 - |  |
| 8 - |  |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |    |                                       |
|----|---------------------------------------|
| 10 | Área de pesquisa                      |
| 7  | Quantidade de periódicos qualificados |
| 1  | Vínculo institucional                 |
| 1  | Possibilidade de interação presencial |

- 1  Se existe a possibilidade de projeto financiado  
10  Bom relacionamento pessoal  
1  Pesquisador PQ/CNPq  
1  Possibilidade de oferta de bolsas  
Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]  
2 -  [área de pesquisa:  
 ]  
3 -  [área de pesquisa:  
 ]  
4 -  [área de pesquisa:  
 ]  
5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Mario Fernando Montenegro Campos (PQ-1B) - **UFMG** - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Uira Kulesza (PQ-2) - **UFRN** - Area: Collaborative and social computing theory, concepts and paradigms

Sim  Não. Comentário:

Wagner Meira Jr. - Area: Spatial-temporal systems

Sim  Não. Comentário:

Antonio Alfredo Ferreira Loureiro (PQ-1A) - **UFMG** - Area: Spatial-temporal systems

Sim  Não. Comentário:

André C. Drummond - **UNB** - Area: Web applications

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

1 -	Software and its engineering
2 -	Computer Systems Organization
3 -	Computing Methodologies
4 -	
5 -	
6 -	
7 -	
8 -	

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

<input type="text" value="10"/>	Área de pesquisa
<input type="text" value="5"/>	Quantidade de periódicos qualificados
<input type="text" value="7"/>	Vínculo institucional
<input type="text" value="4"/>	Possibilidade de interação presencial

- 8  Se existe a possibilidade de projeto financiado
  - 10  Bom relacionamento pessoal
  - 6  Pesquisador PQ/CNPq
  - 6  Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]
- 2 -  [área de pesquisa:  
 ]
- 3 -  [área de pesquisa:  
 ]
- 4 -  [área de pesquisa:  
 ]
- 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Thais Vasconcelos Batista (PQ-2) - **UFRN** - Area: Collaborative and social computing systems and tools

Sim  Não. Comentário:

Uira Kulesza (PQ-2) - **UFRN** - Area: Collaborative and social computing systems and tools

Sim  Não. Comentário:

Luci Pirmez - Area: Distributed architectures

Sim  Não. Comentário:

Flavia Coimbra Delicato - **UFRN** - Area: Distributed architectures

Sim  Não. Comentário:

Anamaria Martins Moreira - **UFRN** - Area: Web applications

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                            |
|-----|----------------------------|
| 1 - | Teoria da Computação       |
| 2 - | Metodologias da Computação |
| 3 - | Computação Aplicada        |
| 4 - | Bioinformática             |
| 5 - | Biologia Computacional     |
| 6 - | Biologia Computacional     |
| 7 - | Biologia Computacional     |
| 8 - | Biologia Computacional     |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="7"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="8"/>  | Vínculo institucional                 |
| <input type="text" value="3"/>  | Possibilidade de interação presencial |

- 4  Se existe a possibilidade de projeto financiado  
 9  Bom relacionamento pessoal  
 1  Pesquisador PQ/CNPq  
 2  Possibilidade de oferta de bolsas  
 Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]  
 2 -  [área de pesquisa:  
 ]  
 3 -  [área de pesquisa:  
 ]  
 4 -  [área de pesquisa:  
 ]  
 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Carlos Hoppen - Area: Symbolic and algebraic algorithms

Sim  Não. Comentário:

Alair Pereira do Lago - <b>USP</b> - Area: Symbolic and algebraic algorithms

Sim  Não. Comentário:

Célia Ghedini Ralha (PQ-2) - <b>UNB</b> - Area: Computational biology

Sim  Não. Comentário:

Ronaldo Fumio Hashimoto (PQ-2) - <b>USP</b> - Area: Computational biology

Sim  Não. Comentário:

Nívio Ziviani (PQ-1A) - <b>UFMG</b> - Area: Database management system engines

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.

# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

- Sim  
 Não  
 Parcialmente. Por que ? Digital Signal Processing né

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                           |
|-----|---------------------------|
| 1 - | Image Processing          |
| 2 - | Computer Vision           |
| 3 - | Digital Signal Processing |
| 4 - |                           |
| 5 - |                           |
| 6 - |                           |
| 7 - |                           |
| 8 - |                           |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="7"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="7"/>  | Vínculo institucional                 |
| <input type="text" value="7"/>  | Possibilidade de interação presencial |

- 7  Se existe a possibilidade de projeto financiado
  - 10  Bom relacionamento pessoal
  - 5  Pesquisador PQ/CNPq
  - 5  Possibilidade de oferta de bolsas
- Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]
- 2 -  [área de pesquisa:  
 ]
- 3 -  [área de pesquisa:  
 ]
- 4 -  [área de pesquisa:  
 ]
- 5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Jussara M. Almeida - Area: Database management system engines

Sim  Não. Comentário:

Alba Cristina Magalhães Alves de Melo (PQ-1D) - <b>UNB</b> - Area: Database management system engines

Sim  Não. Comentário:

Mario Fernando Montenegro Campos (PQ-1B) - <b>UFMG</b> - Area: Database management system engines

Sim  Não. Comentário:

Nivio Ziviani (PQ-1A) - <b>UFMG</b> - Area: Database management system engines

Sim  Não. Comentário:

Marcos André Gonçalves (PQ-1D) - <b>UFMG</b> - Area: Database management system engines

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.



# Questionário de Verificação de Recomendações

1 - Você concorda com o Sistema de Classificação de sub-áreas da Computação criado pela ACM ? O sistema de classificação da ACM foi representado de forma amigável [neste link](#) (clique nos círculos azuis para abrir as sub-áreas). Para mais informações sobre esse sistema de classificação acesse <http://www.acm.org/about/class/2012>.

Sim

Não

Parcialmente. Por que ?

Sugeriria outro sistema de classificação ?

2 - Cite até oito sub-áreas da computação utilizadas em suas pesquisas. Por favor utilize o sistema de classificação da ACM mencionado na questão anterior caso deseje:

- |     |                         |
|-----|-------------------------|
| 1 - | engenharia de software  |
| 2 - | sistemas distribuidos   |
| 3 - | projeto de software     |
| 4 - | arquitetura de software |
| 5 - | testes de software      |
| 6 - | testes de software      |
| 7 - | testes de software      |
| 8 - | testes de software      |

3 - Usando uma escala de 1 a 10 (onde 1 representa total discordância e 10 total concordância), você usaria a base de dados [DBLP](#) como fonte de referência na área de Computação ?

Caso sua avaliação tenha sido inferior a 6, por favor, deixe um comentário abaixo.

4 - Você acredita que ampliar sua rede de colaboração científica com pesquisadores externos a sua instituição tem relação com a avaliação da CAPES do seu programa de pós-graduação ?

5 - Usando uma escala de 1 a 10, sendo 1 o menos importante e 10 o mais importante. Avalie os critérios de escolha de parcerias para publicação:

- |                                 |                                       |
|---------------------------------|---------------------------------------|
| <input type="text" value="10"/> | Área de pesquisa                      |
| <input type="text" value="0"/>  | Quantidade de periódicos qualificados |
| <input type="text" value="0"/>  | Vínculo institucional                 |
| <input type="text" value="0"/>  | Possibilidade de interação presencial |

- 9  Se existe a possibilidade de projeto financiado  
10  Bom relacionamento pessoal  
0  Pesquisador PQ/CNPq  
0  Possibilidade de oferta de bolsas  
Outro, qual ?

[Clique aqui para adicionar critério](#)

6 - Cite até cinco pesquisadores com os quais você teria interesse em publicar (que voce nunca publicou) informando em qual área da Computação se daria essa publicação. Você pode utilizar o sistema de classificação da ACM da questão 1, caso deseje.

- 1 -  [área de pesquisa:  
 ]  
2 -  [área de pesquisa:  
 ]  
3 -  [área de pesquisa:  
 ]  
4 -  [área de pesquisa:  
 ]  
5 -  [área de pesquisa:  
 ]

7 - A metodologia definida neste trabalho identificou as seguintes parcerias para publicação. Por favor, informe se concorda ou não e caso deseje deixar um comentário, utilizar o espaço ao lado. As recomendações foram produzidas automaticamente a partir das áreas utilizadas para publicações em periódicos, segundo o Sistema de Classificação da ACM e não possuem relação de ordenação entre cada uma.

Leonardo Cunha De Miranda - <b>UFRN</b> - Area: Collaborative and social computing systems and tools

Sim  Não. Comentário:

Fabio Kon (PQ-1D) - <b>USP</b> - Area: Collaborative and social computing systems and tools

Sim  Não. Comentário:

Joseph G. Davis - Area: Designing software

Sim  Não. Comentário:

Raphael Cóbe - Area: Designing software

Sim  Não. Comentário:

Ricardo Fernandes - Area: Control path algorithms

Sim  Não. Comentário:

Caso queira deixar um comentário sobre esse trabalho, por favor utilize o espaço abaixo.

Obrigado pela colaboração.