

**ANÁLISE COMPARATIVA DE PESQUISADORES
CONSIDERANDO CARACTERÍSTICAS DE
MÚLTIPLAS ÁREAS DE PESQUISA**

HARLEY AUGUSTO DE LIMA

ANÁLISE COMPARATIVA DE PESQUISADORES
CONSIDERANDO CARACTERÍSTICAS DE
MÚLTIPLAS ÁREAS DE PESQUISA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: MIRELLA MOURA MORO
COORIENTADOR: WAGNER MEIRA JR.

Belo Horizonte

Junho de 2014

© 2014, Harley Augusto de Lima.
Todos os direitos reservados.

L732a Lima, Harley Augusto de
Análise comparativa de pesquisadores considerando
características de múltiplas áreas de pesquisa / Harley
Augusto de Lima. — Belo Horizonte, 2014
xxii, 63 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de
Minas Gerais

Orientador: Mirella Moura Moro
Coorientador: Wagner Meira Jr.

1. Desempenho acadêmico. 2. Índice bibliométrico.
3. Ranking de pesquisador. I. Título.

CDU CDU 519.6*73(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

ANÁLISE COMPARATIVA DE PESQUISADORES CONSIDERANDO CARACTERÍSTICAS DE
MÚLTIPLAS ÁREAS DE PESQUISA

HARLEY AUGUSTO DE LIMA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROFA. MIRELLA MOURA MORO - Orientadora
Departamento de Ciência da Computação - UFMG

PROF. ANA PAULA COUTO DA SILVA
Departamento de Ciência da Computação - UFMG

PROF. CLODOVEU AUGUSTO DAVIS JÚNIOR
Departamento de Ciência da Computação - UFMG

PROF. WAGNER MEIRA JÚNIOR
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 18 de julho de 2014.

Aos meu pais, Maria Augusta e João Augusto.

Agradecimentos

Meus sinceros agradecimentos para todos que me encorajaram e contribuíram de alguma forma para o desenvolvimento deste trabalho. Em especial, gostaria de agradecer:

- A minha orientadora Mirella M. Moro pelos ensinamentos e por acreditar em minha capacidade para desenvolver este trabalho;
- Ao meu coorientador Wagner Meira Jr. por sempre trazer ideias inovadoras;
- Aos professores Alberto H.F Laender e Rodrygo Santos por terem participado de forma ativa neste projeto;
- Aos professores Clodoveu Davis e Ana Paula pelos questionamentos e sugestões;
- Aos meus amigos do LBD e do DCC-UFMG, especialmente ao Thiago Henrique;
- Aos meus pais, pelo apoio, confiança e carinho;
- As minhas irmãs, Haleta Lima e Haline Lima, pelo carinho incondicional que sempre tiveram por mim;
- Aos meu amigos, Afonso Henrique, Felipe Marçal, Henrique Chevreux, Nayara Suzano e Thiago Rafael;
- A Gabriela Flávia, por ter me acompanhado em boa parte dessa jornada;
- Ao Clube Atlético Mineiro, por ter me dado a alegria necessária para conseguir terminar este trabalho;
- A CAPES, CNPq, Fapemig e InWeb por financiar parcialmente este trabalho.

*“Vencer, Vencer, Vencer
Este é o nosso ideal”
(Vicente Motta)*

Resumo

O impacto da produção científica tem sido quantificado tradicionalmente com índices de produtividade, como o popular índice h . Entretanto, diferentes áreas de pesquisas e diferentes subáreas dentro da mesma área podem apresentar padrões de publicação diferentes e que não são avaliados adequadamente utilizando um índice global. Neste trabalho, argumenta-se que índices de produtividade devem considerar peculiaridades de padrões de publicações de diferentes subáreas de pesquisas para produzir uma avaliação imparcial do impacto da pesquisa científica. Inspirado em abordagens de agregação de *ranking*, é proposta uma abordagem para avaliação de pesquisadores com múltiplas subáreas de pesquisa. O método é genérico e produz versões de índices globais comparáveis entre subáreas distintas, como volume de publicação, número de citações e até mesmo do índice h . A avaliação experimental realizada considerando múltiplas subáreas de Ciência da Computação mostra que as versões dos índices comparáveis entre subáreas distintas superam seus correspondentes globais comparados com a avaliação produzida pelo órgão oficial do governo brasileiro. Como resultado, esta dissertação contribui com um mecanismo valioso para apoiar as decisões dos órgãos de financiamento e agências de pesquisa, por exemplo, em uma tarefa de avaliação de pesquisa.

Palavras-chave: Desempenho acadêmico, Índices bibliométricos, Agregação de ranking.

Abstract

The impact of scientific research has traditionally been quantified using productivity indices such as the well-known h-index. On the other hand, different research fields and even different research areas within a single field may have different publishing patterns, which may not be well described by a single, global index. In this work, we argue that productivity indices should account for the singularities of the publication patterns of different research areas, in order to produce an unbiased assessment of the impact of scientific research. Inspired by ranking aggregation approaches in distributed information retrieval, we propose a novel approach for ranking researchers across multiple research areas. Our approach is generic and produces cross-area versions of any global productivity index, such as the volume of publications, citation count and even the h-index. Our thorough evaluation considering multiple areas within the broad field of Computer Science shows that our cross-area indices outperform their global counterparts when assessed against the official ranking produced by the official Brazilian agency. As a result, this thesis contributes a valuable mechanism to support the decisions of funding and research agencies, for example, in any research assessment effort.

Keywords: Research performance, Bibliometric indicators, Ranking aggregation.

Lista de Figuras

1.1	Distribuição de pesquisadores brasileiros em Ciência da Computação por subáreas de interesse. Um pesquisador é considerado de uma subárea se pelo menos 10% das publicações do pesquisador forem classificadas na subárea.	3
3.1	Distribuição dos pesquisadores de Ciência da Computação classificados pelo CNPq, em março de 2012.	15
3.2	Distribuição de tempo desde obtido o título de doutor por categoria do CNPq.	20
3.3	Distribuição do número de alunos de mestrado (a) e doutorado (b) orientados pelas cinco categorias do CNPq.	21
3.4	Distribuição do número de alunos de mestrado (a) e doutorado (b) orientados normalizado pelo tempo de carreira do pesquisador.	21
3.5	Distribuição de (a) volume de publicação e (b) quantidade de citações ao longo das categorias do CNPq.	23
3.6	Número de veículos de publicação por índice h , com o intervalo dos estratos definidos.	25
3.7	Evolução da média por pesquisador de (a) publicação e (b) quantidade de citações nos estratos de índice h ao longo das categorias do CNPq.	26
3.8	Representatividade dos estratos de índice h entre as subáreas	28
3.9	Distribuição de pesquisadores CNPq nas subáreas	29
5.1	Comparação entre o <i>ranking</i> produzido com índice h e com <i>ca</i> -índice.	46
5.2	Distribuição de valores do (a) índice h e (b) <i>ca</i> -índice entre os cinco níveis de relevância considerados.	47
5.3	Comparação entre o <i>ranking</i> de pesquisadores utilizando o volume de publicação e o <i>ca</i> -volume.	48
5.4	Comparação entre o <i>ranking</i> de pesquisadores utilizando o número de citação e o <i>ca</i> -citação.	48

5.5	Distribuição de valores de (a) volume, (b) <i>ca</i> -volume, (c) citação e (d) <i>ca</i> -citação de acordo com os cinco níveis de relevância.	49
5.6	Comparação entre o <i>ranking</i> de pesquisadores com base no índice <i>h</i> , <i>ca</i> -índice e a combinação linear desses.	50
5.7	Pesquisadores promovidos de categoria segundo avaliação do CNPq realizada em 2013.	54

Lista de Tabelas

3.1	Número de veículos de publicação de cada subárea e o resumo das estatísticas (média, desvio padrão e mediana) referente ao volume de publicação e quantidade de citações para os pesquisadores nas 23 subáreas de Ciência da Computação, considerando toda a base de dados construída.	17
4.1	Conjunto de citações dos pesquisadores r_1 , r_2 e r_3	36
4.2	<i>Rankings</i> produzido para cada subárea de acordo com o desempenho dos pesquisadores r_1 , r_2 e r_3	38
5.1	Matriz de confusão para os <i>rankings</i> produzidos pelo <i>ca</i> -índice (colunas) e pela classificação do CNPq (linhas). As células destacadas mostram as discrepâncias entre a classificação dos pesquisadores das categorias 1 e 2. Os valores percentuais foram calculados com relação a <i>todos</i> os pesquisadores considerados.	52
5.2	Matriz de confusão considerando os pesquisadores das categorias 1A-1D em uma ampla categoria 1.	52

Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xix
1 Introdução	1
1.1 Contribuições	3
1.2 Organização da dissertação	4
2 Visão geral sobre avaliação de pesquisadores	5
2.1 Avaliação da produção científica de pesquisadores	5
2.2 Índices bibliométricos	8
2.2.1 Índices bibliométricos com abordagem global	8
2.2.2 Índices bibliométricos com abordagem por área	9
2.3 Considerações finais	11
3 Pesquisadores de Ciência da Computação no Brasil	13
3.1 Conjunto de dados de pesquisadores	14
3.2 Análise dos pesquisadores de Ciência da Computação	18
3.2.1 Tempo de carreira	19
3.2.2 Orientação de alunos	19
3.2.3 Produção científica	22
3.3 Distribuição de qualidade das publicações	22
3.3.1 Classificação de Veículos de Publicação	24
3.3.2 Qualidade das publicações com base nos estratos de índice h	25

3.3.3	Qualificação de publicação por subáreas	27
3.4	Considerações finais	30
4	Ranqueamento de pesquisadores considerando múltiplas subáreas de pesquisa	33
4.1	Premissas de avaliação de produtividade entre subáreas distintas	33
4.2	Método de ranqueamento de pesquisadores entre subáreas distintas . .	35
5	Avaliação experimental	41
5.1	Conjunto de publicações	42
5.2	Rankings de comparação	43
5.3	Procedimento de avaliação	43
5.3.1	Ranking de referência	44
5.3.2	Métrica de avaliação	44
5.4	Resultados	45
5.4.1	Ranking utilizando índice único	45
5.4.2	Ranking utilizando múltiplos índices	50
5.4.3	Análise de falha	51
5.4.4	Análise de promoção dos pesquisadores	53
6	Conclusões e trabalhos futuros	55
	Referências Bibliográficas	59

Capítulo 1

Introdução

Quantificar o desempenho científico de pesquisadores tornou-se um fator importante para desenvolver políticas relacionadas à pesquisa. Estudos quantitativos e qualitativos do desempenho científico provêm uma medida da produção científica, além de servir como estímulo para aumentar a qualidade da pesquisa realizada. Assim, o desempenho do pesquisador medido nesses estudos serve como base de tomadas de decisões relacionadas à pesquisa como promoção, contratação, demissão, distribuição de recursos. Tais análises de desempenho são realizadas por comitês altamente especializados e requerem grande esforço para avaliar a carreira acadêmica dos pesquisadores.

Definir um comitê e realizar avaliações de um conjunto de pesquisadores são tarefas difíceis. Inicialmente, a maioria dos integrantes do comitê devem estar aptos a avaliar grande parte dos pesquisadores candidatos, independente de suas subáreas de interesse. Apesar de vários estudos mostrarem que cada área tem suas próprias características [Bornmann & Marx, 2014; Moksony et al., 2013; Torrisi, 2013], muitas subáreas também possuem características específicas que impactam no padrão de publicação dos pesquisadores [Barbosa & de Souza, 2011]. Além disso, o conjunto de pesquisadores a serem avaliados pode conter alguns pesquisadores concorrendo por apenas uma vaga, ou muitos concorrendo por uma bolsa de pesquisa. Nesse cenário, o problema anterior é ainda pior. Especificamente, se todos os pesquisadores candidatos pertencerem a uma mesma subárea como Recuperação da Informação (RI), essa tarefa é resolvida de forma simples, formando um comitê apenas com especialistas em RI. Entretanto, se os pesquisadores avaliados forem de subáreas distintas dentro da área de Ciência da Computação, pode acontecer de um especialista em Engenharia de Software avaliar um pesquisador de RI. De forma similar, comparar pesquisadores que possuem distintas subáreas de interesse também se mostra uma tarefa complicada e propensa a erros, por exemplo, uma agência de fomento que precise avaliar pesquisadores de diferentes subá-

reas de Ciência da Computação para distribuição de fomento. Considerando que dois pesquisadores de subáreas distintas, um de Teoria Computacional e outro de Computação Gráfica, ambos com o mesmo número de artigos e tendo apenas seus conjuntos de publicações como forma de avaliação: como o comitê deve definir e utilizar um critério (ou um conjunto de) que seja justo suficiente para avaliar tais pesquisadores?

Neste contexto, a decisão de qual pesquisador deve ficar no topo de um *ranking* muitas vezes é baseada em critérios como número de publicações e impacto das publicações. Claramente, a efetividade do *ranking* resultante depende de como e quais critérios são utilizados [Althouse et al., 2009; Oliveira et al., 2012]. Índices bibliométricos têm sido utilizados para medir a produtividade de pesquisadores. Exemplos de tais índices são índice *h* [Hirsch, 2005], índice *g* [Egghe, 2006] e *citation z-score* [Lundberg, 2007]. Do mesmo modo, muitas plataformas acadêmicas virtuais, como ArnetMiner¹, Google Scholar² e Microsoft Academic Search³, utilizam esses índices para produzir *ranking* dos pesquisadores. Entretanto, uma limitação desses índices bibliométricos globais é não considerar a dinamicidade de diferentes subáreas para avaliar o desempenho dos pesquisadores.

Nesse sentido, avaliar pesquisadores sem considerar subáreas de atuação, bem como as especificidades de cada subárea, pode tornar a avaliação injusta. Tais especificidades estão relacionadas a vários aspectos de cada subárea. Por exemplo, a subárea Interação Humano-Computador em Ciência da Computação possui avaliação experimental mais demorada, pois muitas vezes é preciso avaliar o desempenho do usuário diante a utilização de alguma ferramenta [Barbosa & de Souza, 2011]. Enquanto que os experimentos realizados, por exemplo, na subárea de Bancos de Dados muitas vezes são realizados por meio da avaliação de dados gerados de forma automática. Dessa forma, pesquisadores de subáreas distintas possuem padrões de publicação diferentes devido a tais peculiaridades. Para agravar ainda mais esse problema, pesquisadores publicam em mais de uma subárea. Para ilustrar essa observação, a Figura 1.1 mostra a distribuição de pesquisadores brasileiros por subárea em que esses possuem pelo menos 10% de suas publicações⁴. Essa figura mostra uma distribuição normal com uma média de 5 subáreas por pesquisador. Interessante observar que a maioria dos pesquisadores são ativos em mais de 5 subáreas e alguns em até mesmo 13.

Com objetivo de aprimorar tais avaliações de pesquisa acadêmica, nesta dissertação é proposto um método de *ranking* de pesquisadores que pondera múltiplas subáreas

¹ArnetMiner: <http://arnetminer.org/>

²Google Scholar: <http://scholar.google.com/>

³Microsoft Academic Search: <http://academic.research.microsoft.com/>

⁴Mais detalhes sobre a base de dados utilizada para gerar a Figura 1.1 serão apresentados no Capítulo 3.

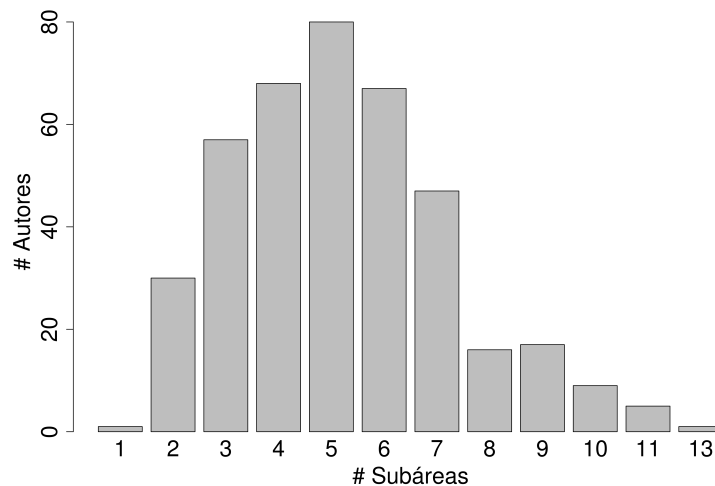


Figura 1.1. Distribuição de pesquisadores brasileiros em Ciência da Computação por subáreas de interesse. Um pesquisador é considerado de uma subárea se pelo menos 10% das publicações do pesquisador forem classificadas na subárea.

de atuação dentro de uma área científica. Em particular, o método estima o desempenho de um pesquisador em cada subárea relativo ao desempenho dos demais pesquisadores na mesma subárea. Assim é garantido que as especificidades da subárea sejam consideradas adequadamente. O desempenho do pesquisador nas múltiplas subáreas é agregado em um *ranking* único projetando o desempenho em cada subárea na subárea base do pesquisador, i.e., subárea em que o pesquisador possui melhor desempenho. O método é genérico e pode ser utilizado para criar versões comparáveis entre subáreas distintas de índices de produtividade globais.

1.1 Contribuições

A contribuição principal desta dissertação é um método de *ranking* de pesquisadores com múltiplas subáreas de pesquisa. Esse método implementa premissas consideradas essenciais para avaliar pesquisadores com múltiplas subáreas de atuação, produzindo versões comparáveis entre subáreas distintas de índices de produtividade globais. Tal método é avaliado de acordo com o *ranking* oficial produzido pelo CNPq⁵ para avaliar pesquisadores de Ciência da Computação.

Para tanto, os seguintes objetivos parciais foram definidos e constituem as contribuições deste trabalho:

- Método para inferir o impacto de veículos de publicação baseado no índice

⁵CNPq: Conselho Nacional de Desenvolvimento Científico e Tecnológico

h [Hirsch, 2005]. Os veículos de publicação são ordenados pelo valor de índice h e, em seguida, são divididos em cinco estratos de impacto.

- Análise dos pesquisadores brasileiros com destacada produção acadêmica na área de Ciência da Computação. Essa análise tem como foco três dimensões: tempo de carreira, alunos orientados, e volume de publicação e quantidade de citações. Além disso, o conjunto de publicações dos pesquisadores é avaliada de acordo com impacto do veículo de publicação e quanto à subárea de atuação do pesquisador.
- Definição de premissas cujo objetivo é proporcionar uma avaliação imparcial de pesquisadores com múltiplas subáreas de interesse. Tais premissas são *Pluralidade* (avaliação deve considerar todas as subáreas de atuação do pesquisador), *Diversidade* (as peculiaridades de cada subárea devem ser consideradas para avaliar o pesquisador) e *Igualdade* (todas as subáreas são igualmente importantes).

1.2 Organização da dissertação

Esta dissertação está organizada da seguinte forma. No Capítulo 2 é apresentada uma visão geral dos estudos de produtividade relacionados a atividades acadêmicas, bem como uma discussão sobre índices de produtividade. No Capítulo 3 é descrita a metodologia empregada no trabalho e é apresentada a análise dos pesquisadores brasileiros de Ciência da Computação com destacada produção acadêmica. No Capítulo 4 as premissas para uma avaliação imparcial de pesquisadores com múltiplas subáreas de atuação são apresentadas, assim como a implementação de tais premissas no método de *ranking* de pesquisadores proposto nesta dissertação. O Capítulo 5 apresenta a avaliação experimental do método de *ranking* proposto. Finalmente, no Capítulo 6 são apresentadas as conclusões e algumas possibilidades para trabalhos futuros.

Capítulo 2

Visão geral sobre avaliação de pesquisadores

Este capítulo discute estudos relacionados à avaliação de pesquisadores, bem como aspectos ligados à bibliometria, amplamente utilizada para avaliar produtividade científica. A Seção 2.1 descreve alguns trabalhos que avaliam a produção científica de pesquisadores e como tais avaliações têm sido realizadas atualmente. A Seção 2.2 apresenta diversos índices bibliométricos, divididos em índices com abordagem global e abordagem por área. Por fim, a Seção 2.3 apresenta algumas considerações finais sobre o enquadramento do trabalho proposto nessa dissertação no contexto de bibliometria.

2.1 Avaliação da produção científica de pesquisadores

Dada a crescente taxa de artigos científicos publicados ao longo dos anos, diversos estudos surgiram para analisar produções científicas com objetivos distintos. Especificamente, estudos comparativos agregam informação, visando descobrir conhecimento a partir de dados de pesquisadores, artigos e citações.

Estudos recentes comparam a pesquisa como um todo no cenário global. Por exemplo, Radicchi & Castellano [2013] analisaram indicadores bibliométricos em uma base de dados com aproximadamente 30 mil pesquisadores de diversas áreas científicas do mundo todo. Tal estudo mostra que o índice h é fortemente correlacionado com o número total de citações recebidas pelo pesquisador e que existe uma correlação entre o índice h e o conjunto de publicações do pesquisador, entretanto tal correlação é muito menos precisa do que a observada para número de citações. Apesar de interessantes,

os resultados são completamente genéricos.

Resultados mais significativos podem ser alcançados tendo uma área científica específica como foco de estudo. Nesse contexto, Kato & Ando [2013] analisaram a produção científica de pesquisadores italianos e concluíram que colaborações internacionais melhoram o desempenho de pesquisadores na área da Química. Os resultados apontaram que a quantidade e a taxa de citação de artigos publicados em veículos locais por pesquisadores com inserção internacional são maiores do que artigos de pesquisadores sem inserção internacional.

Outra perspectiva interessante é centralizar a análise em um determinado país. Sangwal [2013] caracterizou o padrão de publicações, considerando oito países europeus e um sul-americano, o Brasil. Os resultados indicaram que periódicos escritos em inglês possuem fator de impacto maior do que periódicos escritos em outras línguas. Além disso, todos os países analisados possuem periódicos com alto índice de auto citações, mas esse índice varia consideravelmente entre os países. Do mesmo modo, Collazo-Reyes [2014] analisou o impacto de periódicos indexados em países latino-americanos e do Caribe, apontando que o crescimento sem precedente de periódicos no período de 2005-2011 está diretamente relacionado com a alteração da política editorial, e não diz respeito ao crescimento da comunidade científica. De forma semelhante, Michels & Schmoch [2014] concluíram que pesquisadores alemães tendem a publicar em periódicos conhecidos internacionalmente com editora nos Estados Unidos. Vanecek [2014] avaliou o financiamento e os resultados científicos da República Tcheca, mostrando que o número de publicações e patentes cresceram mais rápido do que vários países europeus. Do mesmo modo, Shelton & Lewison [2013] discutiram o cenário da pesquisa na Coreia do Norte, apesar de ser um dos países mais reclusos, mostra-se cada vez mais conectado cientificamente com o mundo, principalmente com a China e a Alemanha.

Em relação ao cenário brasileiro, Glänzel et al. [2006] e Leta et al. [2006] avaliaram o crescimento do Brasil no contexto global. Esses trabalhos apontaram o Brasil como destaque científico na América Latina, ainda que não apresente instituições de pesquisa privadas e concentre o desenvolvimento científico em universidades. Helene & Ribeiro [2011] analisaram a evolução da produção científica considerando o perfil de novos doutores formados, melhoria na infraestrutura e benefícios trazidos com pesquisadores em regime de dedicação exclusiva. Os resultados mostraram que o aumento do número de artigos ocorreu independentemente de investimento em infraestrutura e do aumento de pesquisadores em regime de dedicação exclusiva, sendo que tal crescimento ocorreu, na verdade, em conjunto com aumento do número de alunos de doutorado. Com objetivo de analisar o crescimento de produção de artigos científicos, Almeida & Guimarães [2013] realizaram um estudo sobre artigos publicados por autores brasilei-

ros, comparando os cenários brasileiro e internacional. Esse estudo também identificou áreas científicas que geram conhecimento. De forma geral, todos os artigos mencionados certificam o crescimento científico brasileiro ao longo dos anos.

A respeito de estudos em determinadas áreas científicas, Oliveira et al. [2012] analisaram pesquisadores em Medicina Clínica. Esse trabalho analisou pesquisadores financiados pelo CNPq e compara a produtividade entre as categorias do CNPq utilizando bibliometria e outras variáveis. De acordo com os resultados, melhores indicadores quantitativos e qualitativos são necessários para identificar pesquisadores com destacada produção científica. Roos et al. [2014] conduziram um estudo com foco em Ciências Biológicas, comparando a produção científica de pesquisadores com título de doutor obtido no Brasil e no exterior. Esse trabalho mostrou que artigos de pesquisadores com título de doutor obtido no Brasil foram publicados em periódicos com baixo fator de impacto e receberam menos citações frente aos artigos publicados por pesquisadores que obtiveram título de doutor no exterior.

Entre vários estudos sobre a produtividade de Ciência da Computação no Brasil, Laender et al. [2008] analisaram a qualidade da pesquisa e educação dos melhores programas de pós-graduação em Ciência da Computação no Brasil. Os resultados apontaram a maturidade dos programas de pós-graduação brasileiros comparados a institutos norte-americanos e europeus. De forma similar, Wainer et al. [2009] analisaram a produção científica de Ciência da Computação no Brasil e outros países. Os resultados mostraram que o Brasil tem a maior produção científica entre os países latino-americanos nessa área de conhecimento. Além disso, a produção científica brasileira é cerca de um terço da produção espanhola, um quarto da italiana e aproximadamente a mesma da indiana e russa.

Em relação às subáreas de Ciência da Computação, Arruda et al. [2009] consideraram a distribuição regional do Brasil por tópico de pesquisa, gênero e categorias do CNPq. A respeito da distribuição regional brasileira, existe uma diferença estatística na produção de cada região. Em outro estudo, Wainer et al. [2013] também analisaram a produtividade de subáreas de Ciência da Computação, mostrando que a taxa de publicação e citação diferem entre algumas subáreas, mas não em todas. Além disso, esse estudo mostra que não existe uma correlação significativa entre a taxa de citação e o tamanho da comunidade de uma determinada subárea.

2.2 Índices bibliométricos

Índices bibliométricos têm sido utilizados amplamente pela comunidade científica para mensurar o desempenho de pesquisadores, com objetivo de nortear decisões como critério de recrutamento, progressão de carreira, financiamento de pesquisa, entre outros [Sune Lehmann, 2006]. Idealmente, para mensurar a habilidade do pesquisador em produzir conhecimento científico, duas dimensões devem ser consideradas: quantitativa e qualitativa. Assim, os indicadores comumente utilizados para representar tais dimensões são baseados em número de publicação, quantidade de citações e o impacto do veículo de publicação em que o artigo foi publicado [Torrise, 2013]. No entanto, esses indicadores podem variar de acordo com área científica analisada. A seguir, os índices bibliométricos são discutidos em duas categorias: a Seção 2.2.1 refere-se aos índices com abordagem global, ou seja, índices indiferentes à área de atuação dos pesquisadores avaliados, e a Seção 2.2.2 apresenta os índices que avaliam os pesquisadores considerando possíveis peculiaridades da área de atividade em que o pesquisador está inserido.

2.2.1 Índices bibliométricos com abordagem global

Apesar de o desempenho científico de um pesquisador não poder ser mensurado apenas com índices bibliométricos, esses índices têm sido empregados amplamente para medir a produtividade de pesquisadores. Em plataformas acadêmicas como *Microsoft Academic Search*¹ e *ArnetMiner*², os autores são ordenados de acordo com o índice h [Hirsch, 2005]. O índice h (h é a quantidade de publicações com pelo menos h citações) informa sobre a produtividade e o impacto das publicações do pesquisador em apenas um número. A grande vantagem desse índice é a facilidade para ser calculado. Por outro lado, não é possível comparar o índice h de pesquisadores de áreas de pesquisas e tempo de carreira diferentes, visto que em seu cálculo não é feita nenhuma normalização referente a esses fatores [Bornmann & Marx, 2014].

O índice g [Egghe, 2006] introduz uma melhoria ao índice h . Dado um conjunto de artigos ordenados de forma decrescente de acordo com o número de citações, o índice g é o maior número de tal forma que os g primeiros artigos juntos receberam pelo menos g^2 citações. O índice g tenta amenizar o fato de o índice h ser insensível a artigos altamente citados. Ou seja, uma vez que o artigo pertence ao conjunto h ,

¹*Microsoft Academic Search*: <http://academic.research.microsoft.com/>

²*ArnetMiner*: <http://arnetminer.org/>

não influenciará em nada o valor do índice h do pesquisador se esse artigo duplicar o número de citações.

Tanto índice h quanto o índice g avaliam o desempenho individual do pesquisador sem considerar seus pares. Com o avanço de técnicas de análise de redes sociais [Leskovec et al., 2008; Pallis et al., 2011], trabalhos mais recentes avaliam o pesquisador inserido em sua comunidade. Por exemplo, Freire & Figueiredo [2011] propuseram um índice de produtividade para avaliar pesquisadores em uma rede de colaboração utilizando relacionamento de coautoria. A abordagem possui duas premissas: considera um pequeno grupo de pesquisadores em que avaliação é conduzida e a importância dada para um indivíduo é proporcional à intensidade de suas colaborações com outros indivíduos fora do grupo. Intuitivamente, indivíduos com alto relacionamento fora do grupo servem como pontes onde as ideias e o conhecimento fluem, o que é um aspecto importante para colaboração científica. Nesse mesmo trabalho, os autores também consideram uma variação do índice proposto para avaliar grupos de pesquisa como um todo, avaliando programas de pós-graduação em Ciência da Computação.

Assim como o indivíduo é avaliado considerando o aspecto social do seu grupo, também é possível avaliar o grupo como um todo. Ribas et al. [2013] propuseram uma métrica, denominada *R-Score*, com objetivo de avaliar grupos de pesquisas. A intuição da métrica é que grupos de pesquisa que publicam frequentemente em veículos de publicação em que destacados grupos de pesquisa também publicam, possivelmente são mais produtivos do que grupos que publicam em outros veículos. A métrica foi comparada com o *ranking* de 25 programas de pós-graduação em Ciência da Computação no Brasil, e os resultados indicaram uma boa conformidade entre os *rankings*.

De forma geral, os índices apresentados nessa seção não contextualizam a área em que o pesquisador foca seus trabalhos. Uma avaliação de pesquisadores realizada sem considerar as especificidades de cada área científica é injusta e propensa a erros, como direcionar indevidamente recursos a uma determinada área por essa possuir um volume de publicação maior ou por receber mais citações. Na próxima seção são apresentados os índices que visam capturar a dinâmica da área de atuação do pesquisador para avaliar o impacto de sua produção científica.

2.2.2 Índices bibliométricos com abordagem por área

Um problema relevante na ciência consiste em avaliar pesquisadores de áreas científicas distintas, como Matemática, Física e Ciência da Computação, com base na produção científica [Bornmann et al., 2008; Oliveira et al., 2012; Podlubny, 2005]. Nesse contexto, Podlubny [2005] propôs uma tabela de equivalência de citação para diferentes

áreas científicas, com base na contínua observação da taxa de citação nessas áreas. O trabalho tenta responder se é possível comparar dois pesquisadores de áreas científicas distintas com base na quantidade de citações que esses possuem. Dessa forma, é feita uma normalização para gerar uma tabela de equivalência que estabelece, por exemplo, que uma citação em Matemática corresponde aproximadamente a 5 citações em Engenharia, 19 em Física e 78 em Ciências Biomédicas. Assim, considerar apenas o número total de citações para avaliar pesquisadores de áreas científicas distintas pode ofuscar a quantidade de pesquisadores e o padrão de publicação de cada área. Entretanto, a solução específica proposta em tal estudo requer uma correção periódica da tabela de equivalência, o que pode ser agravada pela evolução instável de cada área de conhecimento.

De forma semelhante, Radicchi et al. [2008] realizaram uma análise sobre a distribuição de citações em publicações de áreas de pesquisas distintas, mostrando que a simples contagem do número citações pode ser uma abordagem enganosa para decidir se um artigo é mais sucedido em comparação com outro de áreas científicas diferentes. Para tal, foi proposto o indicador relativo $c_f = c/c_o$, em que c é a quantidade de citações recebidas por uma publicação e c_o é a média da quantidade de publicações na área de pesquisa em questão. O indicador redimensiona a distribuição de citação para os artigos de diferentes áreas científicas sobre a mesma inclinação quando c_f é aplicado. Não importa as áreas científicas, Biologia, Física Nuclear ou Engenharia Espacial, a distribuição do fator c_f é a mesma para todas as áreas. Além disso, esse indicador c_f permite comparar devidamente, dentro de uma mesma área científica, artigos publicados em anos diferentes. Posteriormente, Bornmann & Daniel [2009] avaliaram as vantagens desse indicador utilizando uma base de dados da área de Química, comparando suas subáreas. Os resultados desse estudo mostraram que outras métricas em algumas situações atingem um resultado mais adequado para produzir uma distribuição universal de citação nessa área específica de conhecimento.

Claro & Costa [2011] propuseram o *x-index* como um índice bibliométrico para comparar pesquisadores de áreas de pesquisa distintas. O índice considera os melhores autores de cada área de conhecimento (com base no volume de publicação) como um conjunto de referência. Assim, a produtividade dos demais autores é calculada relativa ao conjunto de referência da área. Além disso, o índice visa atingir três objetivos: refletir tanto a quantidade quanto a qualidade das publicações, demandar moderada extração de dados e esforço de processamento, e permitir comparar perfis de áreas científicas distintas. Apesar desse índice apresentar vários méritos, também possui limitações, como a vulnerabilidade em questão de escala. Ou seja, é possível que pesquisadores com alta taxa de publicação em periódicos com alto fator de impacto te-

tenham valores de *x-index* similares aos pesquisadores com baixa taxa de publicações em periódicos com baixo fator de impacto. Ademais, os autores escolhidos como conjunto de referências podem representar valores discrepantes, visto que volume de publicação não é necessariamente um indicador de qualidade.

Outro índice bibliométrico que visa reduzir possíveis discrepâncias para avaliar pesquisadores de áreas de pesquisas distintas é o *crown-index* [Waltman et al., 2010]. Em particular, tal índice é construído com base na premissa que a média do número de citações por publicação entre as distintas áreas de pesquisa varia [Leydesdorff, 2012; Podlubny, 2005]. Com base nessa premissa, dado um conjunto de publicações, o *crown-index* normaliza a quantidade de citações de um pesquisador pelo número esperado de citações em cada área científica. O número esperado de citações de uma publicação é igual à média de citações de publicações do mesmo tipo (i.e artigo, cartas ou artigo de revisão) publicadas na mesma área científica e no mesmo ano [Waltman et al., 2011].

A normalização da taxa de citação realizada no cálculo do *crown-index* não é feita individualmente para cada publicação, mas em um nível mais elevado de agregação (i.e pesquisador, grupo ou departamento de pesquisa). Assim, a taxa de citação média de um pesquisador é comparada com a taxa de citação média da área de pesquisa em que o pesquisador publica. Dessa forma, a normalização favorece publicações antigas de áreas científicas com alto índice de citação. Para contornar esse impasse, Lundberg [2007] propôs melhorias no mecanismo de normalização do *crown-index*, criando o *citation z-score*. No *citation z-score*, a normalização é feita no nível da publicação, atribuindo pesos às publicações de acordo com a distribuição de citações nos artigos.

2.3 Considerações finais

Como mostrado na Seção 2.1, estudos que avaliam a produção científica de pesquisadores possuem pontos em comuns para nortear as análises. Vários estudos têm como foco uma área científica específica como Química, Matemática, Ciência da Computação [Wainer et al., 2009; Oliveira et al., 2012; Kato & Ando, 2013]. Além disso, vários estudos analisam a produção científica em um determinado país, mostrando uma visão geral de sua produção e em várias situações comparando com o contexto mundial [Lander et al., 2008; Sangwal, 2013; Collazo-Reyes, 2014; Michels & Schmoch, 2014; Vanecek, 2014]. Nesses estudos, vários aspectos da carreira acadêmica do pesquisador são avaliados, como conjunto de publicações, quantidade de citações, colaborações internacionais, etc. Torrisi [2013] argumenta que, dada a complexidade da produção acadêmica, a produtividade do pesquisador deve ser relacionada a quatro dimensões:

publicação científica, premiação e atividades acadêmicas (editor, coordenador de projetos, etc), patentes e atividades didáticas.

Nesse sentido, esta dissertação inicia com uma avaliação dos pesquisadores brasileiros de Ciência da Computação com destacada produção acadêmica com base em três critérios: tempo de carreira (tempo desde a obtenção do título de doutor), orientação de alunos (quantidade de orientações concluídas de alunos de mestrado e doutorado) e produção científica (número de artigos e quantidade de citações recebidas). Diferentemente dos demais estudos, nesta análise, a produção científica é investigada com ênfase na qualidade do veículo de publicação e na subárea de trabalho do pesquisador. De forma geral, essa análise mostra que o padrão de publicação de cada subárea influencia no padrão de publicação do pesquisador, o que motivou a implementação de um método que considere tais diferenças.

A maioria dos índices apresentados na Seção 2.2 avaliam os pesquisadores utilizando indicadores globais, como volume de publicações e quantidade de citações. Conforme discutido em [Waltman et al., 2010; Claro & Costa, 2011], esses indicadores variam consideravelmente entre áreas distintas, até mesmo entre subáreas. Assim, em uma avaliação com pesquisadores com diversas subáreas de atuação, tais indicadores podem levar a interpretações equivocadas. Então, vários trabalhos surgiram propondo abordagens que permitam a comparação de pesquisadores de áreas distintas [Podlubny, 2005; Lundberg, 2007; Radicchi et al., 2008; Bornmann & Daniel, 2009]. Entretanto, tais métricas falham em capturar o desempenho dos pesquisadores entre as diversas subáreas de atuação do pesquisador dentro de uma área científica.

Apesar dos recentes esforços para criar métricas que possibilitem a comparação de pesquisadores de diferentes subáreas, o método proposto nesse trabalho é a primeira tentativa de produzir um *ranking* com base no desempenho dos pesquisadores nas diferentes subáreas de um mesma área científica. Especificamente, a abordagem proposta é capaz de criar versões de critérios de *ranking* para avaliar pesquisadores de diferentes subáreas, com base em todas as subáreas de atuação do pesquisador. Portanto, o método normaliza índices como volume de publicações, quantidade de citações, tornando viável a comparação desses índices entre subáreas.

Capítulo 3

Pesquisadores de Ciência da Computação no Brasil

Quantificar o desempenho de um pesquisador individualmente ou de um grupo de pesquisadores proporciona uma medida relativa da produção científica e representa um estímulo para melhoria da qualidade da pesquisa realizada. Tais análises tornam-se ainda mais úteis quando os resultados auxiliam o governo ou agências de financiamento em relação a novas direções a serem tomadas relativas ao financiamento ou quanto as políticas de pesquisa. Na verdade, análises quantitativas e qualitativas da produção acadêmica de pesquisadores são importantes para os governos alcançarem melhores resultados [Lamont, 2012] e para as universidades buscarem melhores posições em *ranking* globais [Torrise, 2013]. Além disso, avaliações de pesquisadores podem identificar centros de excelência, que podem liderar as ações relacionadas a área de pesquisa em questão, ou centros que necessitam de investimento [Abramo et al., 2011]. Assim, tanto na academia quanto na indústria, importantes decisões relacionadas à carreira do pesquisador dependem do desempenho do mesmo.

Neste capítulo é feita uma análise para abordar tais questões avaliando o desempenho dos pesquisadores brasileiros de Ciência da Computação considerando três dimensões centrais: tempo de carreira, quantidade de alunos orientados, e volume de publicação e quantidade de citações. Em seguida, é feita uma análise aprofundada referente às publicações dos pesquisadores. Nesse sentido, é proposto um método para medir o impacto de veículos de publicação baseado no índice h , assim as publicações são avaliadas de acordo com tal método e a subárea de interesse dos pesquisadores.

A seguir, a Seção 3.1 apresenta a metodologia empregada nas análises mencionadas. A seção 3.2 apresenta a avaliação do perfil dos pesquisadores, enquanto que a Seção 3.3 apresenta o método proposto para quantificar o impacto dos veículos de

publicação e faz uma análise mais aprofundada do conjunto de publicações dos pesquisadores.

3.1 Conjunto de dados de pesquisadores

Tanto na área acadêmica quanto na indústria, a carreira de um pesquisador está sob constante avaliação para diferentes propósitos, incluindo promoção, realocação, premiação, distribuição de recursos e até mesmo para alocação de alunos. A análise de desempenho do pesquisador normalmente depende de um comitê altamente especializado, que deve se reunir, definir um conjunto de critérios de avaliação e, por fim, aferir o desempenho dos pesquisadores. Tal processo pode ser bem dispendioso em termos de tempo, visto que analisar um número alto de pesquisadores não é uma tarefa simples. Para as análises propostas neste trabalho, é necessário definir um conjunto de pesquisadores e coletar seus dados para então serem avaliados. A seguir, são descritos os quatro passos para construção da base de dados utilizada neste trabalho.

Passo 1: Definição do conjunto de pesquisadores. Atualmente no Brasil, existem aproximadamente 2.700 programas de pós-graduação em 46 áreas científicas, incluindo Antropologia, Biologia, Economia, Ciência da Computação, Engenharia, Medicina, Filosofia, entre outras. Especificamente em Ciência da Computação, existem 69 programas de pós-graduação espalhados por todo o país¹. Considerando todos os potenciais pesquisadores para o presente estudo, a construção da base de dados teve como foco pesquisadores de Ciência da Computação que fazem parte do Programa de Bolsa de Produtividade do CNPq. Esse programa financia os pesquisadores mais produtivos em cada área científica. Para tanto, o CNPq conta com 46 comitês altamente especializados que avaliam todos os candidatos e os classificam em cinco categorias: *1A*, *1B*, *1C*, *1D* e *2*, em ordem decrescente de valor de bolsa e taxa de bancada.

Para classificar os pesquisadores em categorias, cada comitê avalia um projeto de pesquisa (escrito pelo pesquisador candidato) seguindo um conjunto de critérios² e o perfil acadêmico nos últimos cinco anos. O perfil acadêmico do pesquisador abrange: publicação em periódicos, conferências e livros; orientação de alunos; contribuição para ciência, tecnologia e inovação (incluindo patentes); coordenação e participação de projetos de pesquisa; inserção internacional; participação como editor científico; entre outras atividades no âmbito acadêmico.

¹De acordo com CAPES 2013 em <http://avaliacaotrienal2013.capes.gov.br>

²Critérios de avaliação utilizados pelo CNPq: <http://www.cnpq.br/web/guest/criterios-de-julgamento>

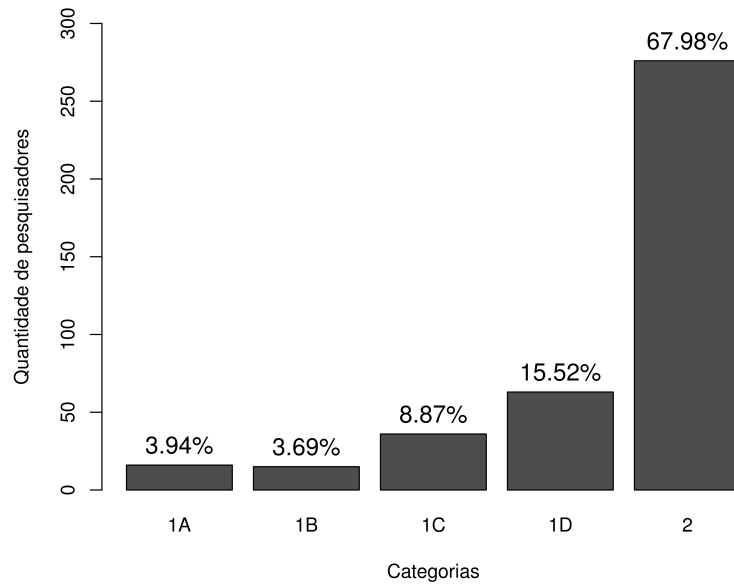


Figura 3.1. Distribuição dos pesquisadores de Ciência da Computação classificados pelo CNPq, em março de 2012.

Em Ciência da Computação, existem 406 pesquisadores distribuídos nas cinco categorias do CNPq, conforme mostrado na Figura 3.1. Considerando a edição de 2011 (resultados publicados em Março de 2012), 32% dos pesquisadores estão no nível 1 e o 68% restante no nível 2. O ponto de entrada no programa de financiamento é a categoria 2, que concentra a maioria dos pesquisadores uma vez que o valor da taxa de bancada é menor do que a recebida pelos pesquisadores da categoria 1.

Passo 2: Coleta dos dados dos pesquisadores. Definida a lista dos 406 pesquisadores de Ciência da Computação, o próximo passo é coletar os dados desses pesquisadores. Todos candidatos do programa do CNPq devem informar o histórico acadêmico na plataforma Lattes³, uma das mais proeminentes iniciativas para arquivamento de atividades acadêmicas [Lane, 2010]. Em linhas gerais, essa plataforma disponibiliza uma ferramenta para adicionar e publicar *curriculum vitae* de pesquisadores. Dentre as informações disponíveis estão: formação acadêmica, atividades profissionais, produção científica, orientação de alunos, premiação, participação em eventos, entre outras. Nesse passo, foi coletada toda informação necessária dos 406 pesquisadores considerados nesse trabalho.

Passo 3: Processo de tratamento dos dados dos pesquisadores. Os dados coletados do Lattes precisaram passar por um processo de limpeza, sobretudo os dados

³Plataforma Lattes: <http://lattes.cnpq.br>

relacionados às publicações. Especificamente, para cada pesquisador, o conjunto de publicações pode ser atribuído indevidamente a outro autor, se diferentes autores possuem a mesma variação de nomes, ou pode ser dividido em grupos associados a diferentes variações do nome do pesquisador. Assim, publicações de um autor podem não estar agrupadas adequadamente devido à ambiguidade de nomes: o mesmo autor pode ter variações de nomes (sinônimos), ou autores distintos podem ter nomes similares (polissemia) [Ferreira et al., 2012], levando à separação ou ao agrupamento errôneo de citações [Lee et al., 2007]. Para resolver a ambiguidade dos nomes de autores, foi aplicado o método estado-da-arte proposto por Cota et al. [2010]. Esse método gera os grupos iniciais de publicações de acordo com a similaridade das publicações, como título e veículo de publicação, e a união desses grupos é realizada considerando a rede de coautoria dos pesquisadores.

Passo 4: Classificação de artigos em subáreas. Como será apresentado nas próximas seções, uma informação importante neste trabalho é a subárea de cada publicação. Para tanto, o passo de classificação em subáreas é dividido em duas etapas: uma para os artigos publicados em conferências e a outra para artigos publicados em periódicos, como explicado a seguir. É importante observar que considerar apenas publicações de periódicos não é suficiente para Ciência da Computação. Assim como os pesquisadores de sociologia devem considerar livros em uma avaliação de desempenho acadêmico [Moksony et al., 2013], cientistas da computação tendem a publicar seus resultados inovadores principalmente em conferências [Laender et al., 2008], não em periódicos.

Para publicações em conferências, foram obtidos os dados do projeto SHINE⁴ (*Simple H-Index Estimator*). Esse projeto reúne publicações de conferências com base em uma lista de veículos de publicação elaborada pelas Comissões Especiais da Sociedade Brasileira de Computação (SBC). Existem atualmente 23 comissões destinadas às subáreas de Ciência da Computação, tais com Bancos de Dados, Interação Humano-Computador, Inteligência Artificial, entre outras. Neste trabalho, foi utilizada a base de dados do SHINE construída em 2011 que contém mais de 800.000 publicações de aproximadamente 1.800 conferências e 7,5 milhões de citações coletadas no começo de 2011. Por fim, cada conferência pode ser classificada em mais de uma subárea, de acordo com a lista sugerida pelas Comissões Especiais.

Para publicações em periódicos, foi considerada a lista de periódicos classificados

⁴SHINE: <http://shine.i.comp.ufam.edu.br>

Subárea	#Conf.	#Per.	Volume			Citação		
			Med.	σ	Mediana	Med.	σ	Mediana
Bancos de Dados	184	127	8,55	13,99	4	73,3	182,51	12
Biologia Computacional	25	28	2,55	2,53	2	14,73	19,17	7
Comp. Aplicada a Saúde	25	67	3,56	5,83	2	13,7	34,21	3
Computação Gráfica ^a	108	105	9,95	10,97	5	58,34	115,63	13
Computação Musical	15	6	2,59	3,28	2	4,21	8,21	0
Concep. de Circuitos ^c	124	112	5,61	12,36	2	34,93	122,71	5
Engenharia de Software	95	42	7,72	12,71	3	57,98	144,89	13
Geoinformática	14	11	4,59	7,92	2	20,95	42,42	2
Informática na Educação	35	37	3,25	5,04	1	10,49	23,45	1
Inteligência Artificial	264	163	9,82	13,24	5	72,56	119,9	26
Interação Humano Computador	21	31	2,71	3,19	1	17,48	26,15	6
Jogos e Entretenimento	17	6	2,37	3,06	1	13,97	56,92	0
Linguagens de Programação	56	23	3,81	4,46	2	47,23	104,88	8
Métodos Formais	49	68	2,67	4,13	1	17,51	32,93	4
Process. de Linguagem Natural	59	43	4,26	4,41	2	37,8	95,34	7
Redes de Comp. ^b	297	161	13,46	18,77	6	84,95	177,22	29
Redes Neurais	84	82	7,83	12,1	4	40,52	77,37	14
Robótica	56	63	3,24	3,8	2	29,71	70,27	4
Segurança de Sist. Comp.	100	98	10,05	12,81	4	31,61	142,57	3
Sist. Tolerantes a Falhas	32	7	2,45	3,02	1	23,35	54,8	5
Sistemas Colaborativos	10	14	8,25	16,02	2	49	101,16	6
Sistemas de Informação	487	188	22,02	19,15	17	160,01	230,76	75,5
Teoria Computacional e Alg.	354	188	10,65	9,97	8	95,98	151,95	41

^a Computação Gráfica e Processamento de Imagem

^b Redes de Computadores e Sistemas Distribuídos

^c Concepção de Circuitos e Sistemas Integrados

Tabela 3.1. Número de veículos de publicação de cada subárea e o resumo das estatísticas (média, desvio padrão e mediana) referente ao volume de publicação e quantidade de citações para os pesquisadores nas 23 subáreas de Ciência da Computação, considerando toda a base de dados construída.

pelo Qualis⁵. O Qualis é uma iniciativa da CAPES⁶ (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*) que classifica periódicos nos quais pesquisadores brasileiros tenham publicado nos últimos três anos. Para obter a referência completa dos artigos, a lista de periódicos foi confrontada com XML da DBLP⁷, utilizando o *International Standard Serial Number* (ISSN) ou, na ausência do ISSN, o título do periódico. Com isso, a base de dados proveniente do SHINE foi expandida com 271.000 artigos de 188 periódicos. Dado que cada artigo de conferência é classificado em subáreas (de acordo com as Comissões Especiais), os artigos de periódicos também precisam ser classificados nessas mesmas 23 subáreas. Como um periódico pode ser classificado em mais de uma subárea, foi utilizado a versão de multi-classe do algoritmo de classificação *Lazy Associative Classifier* (LAC) [Velooso et al., 2007]. Para tanto, foram utilizados

⁵Qualis: <http://qualis.capes.gov.br>

⁶CAPES: <http://www.capes.gov.br/>

⁷DBLP: <http://dblp.uni-trier.de/>

os artigos de conferências, previamente classificados em subáreas, como conjunto de treino e os artigos de periódicos como conjunto de teste.

No total, foram consideradas publicações de 2001 a 2011, classificadas em 23 subáreas de Ciência da Computação. A cobertura média das 23 subáreas, em relação aos artigos publicados em conferências, é de aproximadamente 88%. Para 19 subáreas, a base do SHINE cobre mais de 80% da lista de referência de conferências fornecida pelas Comissões Especiais da SBC. Para 13 subáreas, a cobertura é superior a 90%, e apenas para uma subárea a cobertura é menor que 70%. A Tabela 3.1 mostra mais detalhes sobre toda a base de dados gerada.

3.2 Análise dos pesquisadores de Ciência da Computação

Definida a base de dados com toda informação relevante dos pesquisadores e suas publicações, nesta seção é apresentada a caracterização do perfil dos pesquisadores brasileiros de Ciência da Computação. Como mostrado por estudos recentes, volume de publicações e quantidade de citações devem ser analisados sob a perspectiva da experiência de carreira de cada pesquisador. Especificamente, duas variáveis relevantes são fáceis de serem definidas: tempo de carreira [Sugimoto & Cronin, 2012] e alunos orientados [Kutlar et al., 2013; Miller et al., 2013; Torrisi, 2013]. Na maioria dos casos, a carreira profissional de um pesquisador começa depois de adquirido o título de doutor. Com o passar do tempo, o pesquisador estabelece colaborações que vão além do trabalho relacionado com seu orientador, aumentando o número de publicações e citações [Ibáñez et al., 2013]. Ademais, no âmbito acadêmico, além das colaborações estabelecidas, o pesquisador pode contar com os trabalhos dos alunos que estão sob sua orientação. Normalmente, quanto mais alunos um pesquisador orienta, mais trabalhos são publicados. Apesar de estudos anteriores focarem em alunos de doutorado [Kutlar et al., 2013], no presente trabalho são considerados tanto alunos de doutorado quanto de mestrado, pois dos 69 programas de pós-graduação no Brasil, apenas 25 concedem título de doutor. Em geral, comparar volume de publicações e citações de um recém doutor com um professor com uma longa carreira profissional pode gerar resultados tendenciosos. Assim, a análise proposta tem como objetivo responder as seguintes questões:

Q1: Como o tempo de carreira (anos desde que foi adquirido o título de doutor) pode impactar a produtividade do pesquisador?

Q2: Como o número de alunos orientados, tanto de mestrado quanto de doutorado, pode impactar na produtividade do pesquisador?

Q3: Como o volume de publicações e o número de citações influenciam na posição do pesquisador no *ranking*?

A seguir, os pesquisadores são analisados em termos de tempo de carreira (Seção 3.2.1), alunos orientados (Seção 3.2.2) e produção científica (Seção 3.2.3).

3.2.1 Tempo de carreira

Para responder a questão de pesquisa Q1, é considerado o ano em que cada pesquisador obteve o título de doutor. Dado os pesquisadores agrupados nas categorias do CNPq, a Figura 3.2 mostra a distribuição de tempo desde que os pesquisadores receberam o título de doutor. Note que os valores da mediana decaem da categoria 1A para categoria 2. De forma geral, quanto mais experiente é o pesquisador (experiência em termos de tempo de doutorado), melhor sua posição no *ranking* do CNPq. Entretanto, existem sete pesquisadores com tempo de carreira discrepantes na categoria 2, o que mostra que é necessário mais do que experiência (em termos de tempo de carreira) para subir de categoria.

É possível notar também que existem dois intervalos maiores entre as medianas da categoria 1A para 1B (6,5 anos) e 1D para 2 (7 anos), enquanto os intervalos entre as medianas das categorias 1B para 1C (2,5 anos) e 1C para 1D (1,5 anos) são menores. Diante de uma análise mais detalhada, esses obstáculos para mudar da categoria 1B para 1A e da 2 para 1D podem ser parcialmente explicados com as regras do CNPq para promover pesquisadores. Por exemplo, um pesquisador para ser promovido ao nível 1 deve possuir pelo menos oito anos de tempo de carreira. Além disso, para pertencer à categoria 1A, o pesquisador deve ter contribuições claras para as comunidades nacional e internacional, que demanda tempo e esforço para serem alcançados.

3.2.2 Orientação de alunos

Esta seção aborda a questão de pesquisa Q2, comparando o número de alunos de mestrado e doutorado orientados pelos pesquisadores. A Figura 3.3 mostra a distribuição em cada categoria do CNPq. Como pode ser visto, exceto entre as categorias 1A e 1B, existe um decaimento claro da categoria 1B para categoria 2 em ambos os gráficos. Na Figura 3.3 (a), o número de alunos de mestrado orientados pelos pesquisadores da categoria 1A e 1B são similares, e os valores das medianas são 34, 35, 25, 20 e

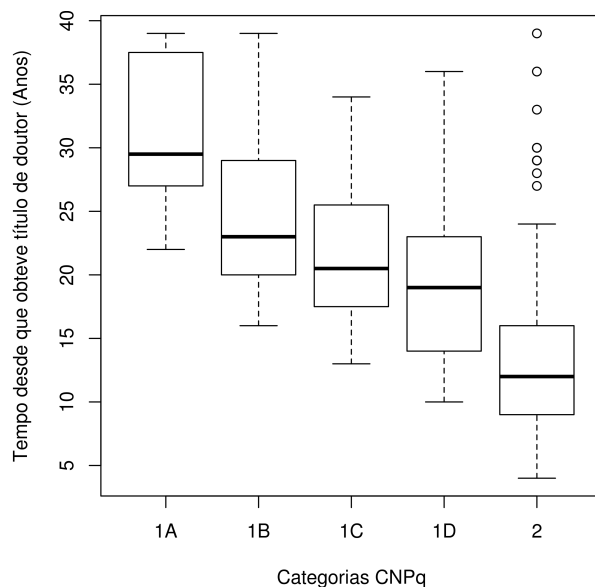


Figura 3.2. Distribuição de tempo desde obtido o título de doutor por categoria do CNPq.

11, respectivamente da categoria *1A* para *2*. Na Figura 3.3 (b), a diferença entre as medianas da categoria *1A* e *1B* é notável, bem como a igualdade entre *1A* e *1C*. Os valores das medianas referentes aos alunos de doutorado orientados são 9, 12, 9, 5 e 1, respectivamente da categoria *1A* para *2*.

Ainda de acordo com a Figura 3.3, a categoria *1A* possui a mediana igual ou menor que a da categoria *1B*. Em particular, a mediana do número de alunos de doutorado orientados da categoria *1A* é igual a da categoria *1C*. O desempenho da categoria *1A* pode ser explicado pelas atividades administrativas que os pesquisadores dessa categoria normalmente assumem (reitor, pró-reitor, diretor, membro de comissão de agência de fomento, etc). Assim, a orientação de alunos não é a única prioridade dos pesquisadores que estão nas categorias superiores.

Para eliminar qualquer viés na análise anterior devido ao tempo de carreira de cada pesquisador, na Figura 3.4 a quantidade de alunos orientados é normalizada pelo tempo de carreira do pesquisador. A proporção de alunos de mestrado orientados é bastante similar entre as categorias do CNPq, com o valor da mediana perto de um aluno por ano de carreira. Assim, esses resultados não podem ser utilizados para distinguir pesquisadores de categorias distintas. Por outro lado, a proporção de alunos de doutorado varia ao longo das categorias do CNPq. O valor da mediana decresce da categoria *1B* para categoria *2*, e o valor da mediana da categoria *1A* é bastante similar ao da categoria *1D*, perto de um aluno de doutorado em cada dois anos de tempo de carreira.

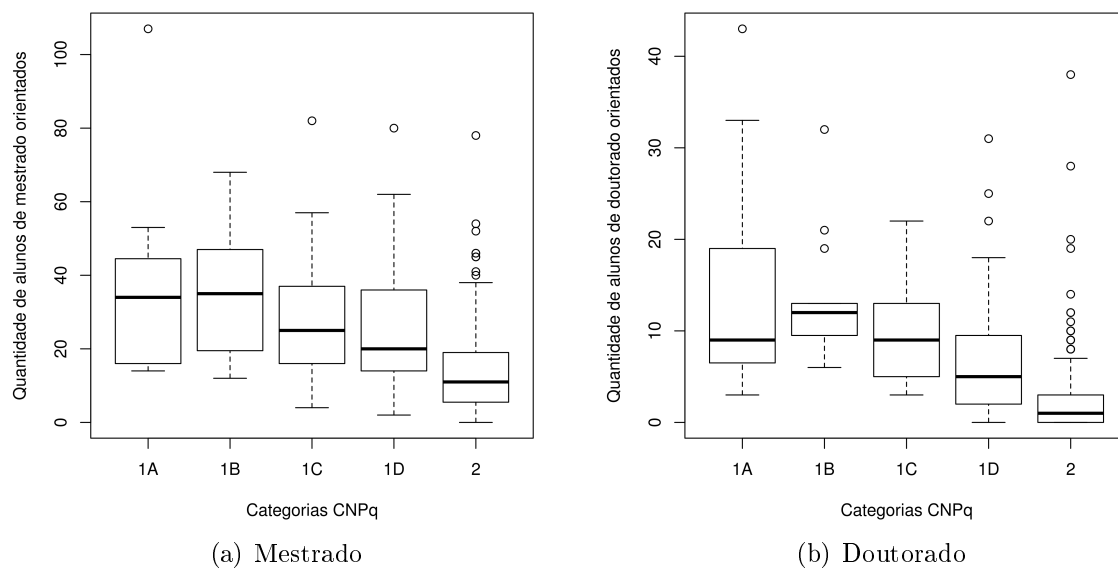


Figura 3.3. Distribuição do número de alunos de mestrado (a) e doutorado (b) orientados pelas cinco categorias do CNPq.

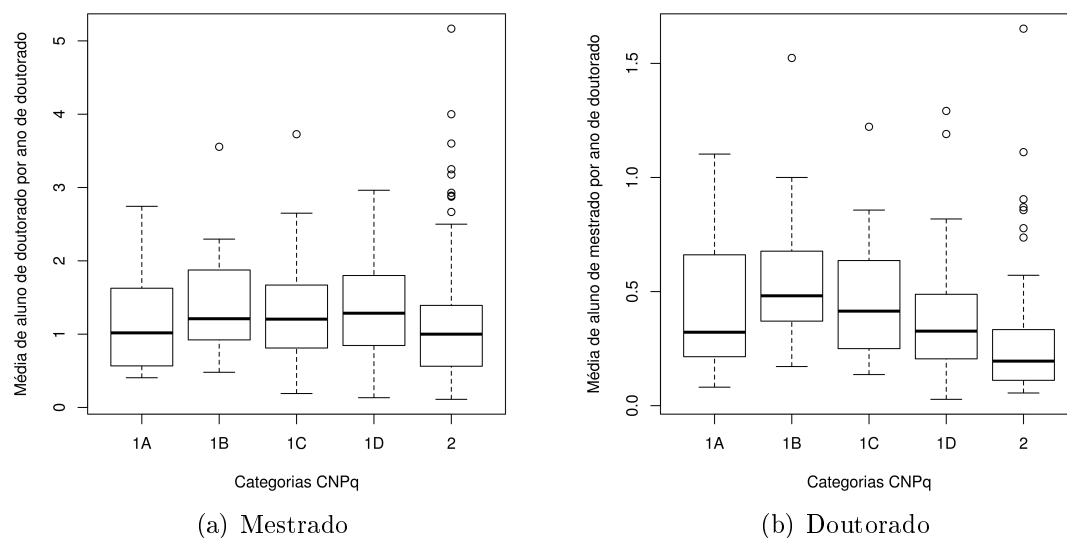


Figura 3.4. Distribuição do número de alunos de mestrado (a) e doutorado (b) orientados normalizado pelo tempo de carreira do pesquisador.

É importante notar também que as Figuras 3.3 e 3.4 apresentam vários pesquisadores com valores discrepantes na categoria 2. Afim de explicar tais valores, considere também a Figura 3.2. Essa figura mostra pesquisadores com mais de 25 anos de carreira na categoria 2. Conseqüentemente, esses pesquisadores orientaram mais alunos do que os demais nessa categoria, o que explica os valores discrepantes nas Figuras 3.3 e 3.4

3.2.3 Produção científica

Nessa seção, a produção científica dos pesquisadores é analisada com base em suas publicações. Para responder a questão de pesquisa Q3, os pesquisadores são comparados em termos de volume de publicação e número de citações ao longo das categorias do CNPq. Lembrando que são consideradas somente as publicações de 2001 a 2011 em todas análises realizadas nesse trabalho.

A Figura 3.5 apresenta a distribuição de publicações em cada categoria do CNPq, considerando (a) número de publicações e (b) quantidade de citações. Note que cada ponto representa um pesquisador, e o valor atribuído a cada pesquisador é normalizado pelo maior valor, afim de limitar os valores entre 0 e 1. Em termos de volume de publicação na Figura 3.5 (a), o valor da mediana é descendente da categoria *1B* à categoria *2*, entretanto o valor da mediana da categoria *1A* é menor do que a da categoria *1B* e possui valor similar ao da categoria *1C*. Uma possível explicação para alta produtividade da categoria *1B* é o alto número de alunos orientados por essa categoria, como mostrado na Figura 3.3. De forma semelhante, o desempenho alcançado pelos pesquisadores da categoria *1A* pode estar relacionado com as atividades administrativas que esses pesquisadores devem realizar. O mesmo padrão pode ser visto em termos da distribuição da quantidade de citações de acordo com as categorias do CNPq. Novamente, o valor da mediana decresce da categoria *1B* para categoria *2*, e o valor da mediana da categoria *1A* é menor do que a mediana da categoria *1B*. Apesar da categoria *1B* possuir valor da mediana maior, apenas um pesquisador conseguiu uma pontuação maior que 0,4, enquanto que na categoria *1A* cinco pesquisadores conseguiram pontuação superior a 0,4.

De forma geral, a distribuição de volume de publicações e quantidade de citações entre as categorias do CNPq seguem a distribuição esperada (i.e. valor da mediana decresce ao longo das categorias), exceto a categoria *1A*. Como mencionado anteriormente, para um pesquisador ser promovido para as categorias superiores, toda carreira acadêmica deve ser considerada, não apenas o conjunto de publicações.

3.3 Distribuição de qualidade das publicações

Na seção anterior é feita a caracterização do perfil dos pesquisadores brasileiros de Ciência da Computação. Nesta seção é feita uma análise mais aprofundada de suas publicações. Especificamente, uma avaliação de um grupo de pesquisadores deve considerar o perfil acadêmico do pesquisador como um todo [Sugimoto & Cronin, 2012]. No entanto, o conjunto de publicações geralmente é o fator mais importante para avaliar

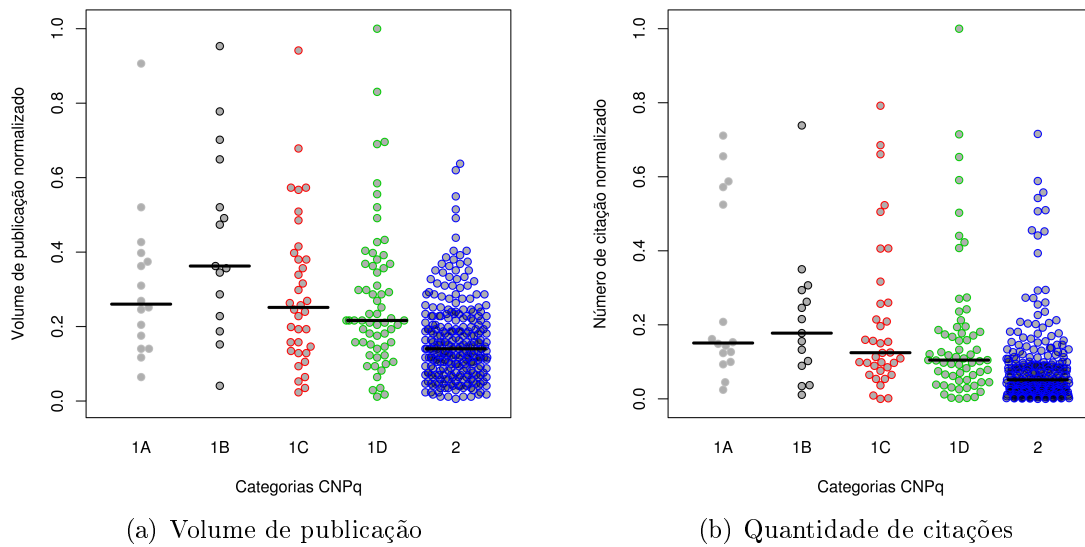


Figura 3.5. Distribuição de (a) volume de publicação e (b) quantidade de citações ao longo das categorias do CNPq.

os pesquisadores. Nesse contexto, definir um critério de qualidade para veículos de publicação pode reduzir o desbalanceamento nessas avaliações. Note que simplesmente contar o número de publicações de um pesquisador não é suficiente, pois é necessário considerar também a qualidade do veículo de publicação em que o artigo é publicado. Essa é uma maneira de se evitar que um pesquisador com várias publicações seja melhor avaliado do que um pesquisador com menos artigos, mas em veículos de publicação com maior impacto.

A maioria dos estudos de avaliação de produtividade de pesquisa têm como foco o serviço de indexação *Thomson Reuters Web of Science*⁸ (WoS), que utiliza o Fator de Impacto para avaliar os veículos de publicação. Especificamente para a comunidade de Ciência da Computação, o Fator de Impacto recebe várias críticas devido a três fatores. Primeiro, a maioria das publicações de Ciência da Computação é direcionada para conferências, e não periódicos [Laender et al., 2008]. O serviço de indexação WoS não inclui anais de congressos, exceto aqueles publicados na *Springer's Lecture Notes in Computer Science Series* (LNCS) e em conferência ASIST. Além disso, alguns periódicos com forte tradição em Ciência da Computação não são indexados pelo WoS, como *ACM Journal on Experimental Algorithms*, *ACM Transactions on Algorithms*, *Journal of Discrete Algorithms*, entre outros. Por fim, o WoS inclui alguns periódicos que não são considerados pela comunidade de Ciência da Computação, como *Journal of Chemical Information and Modeling* [Wainer et al., 2009, 2011].

⁸Thomson Reuters Web of Science: <http://thomsonreuters.com/thomson-reuters-web-of-science/>

Como problemas similares ocorrem em outras áreas – e.g., em Sociologia [Moksony et al., 2013], a agência brasileira responsável por avaliar programas de pós-graduação (CAPES) desenvolveu seu próprio mecanismo de avaliação. Especificamente, o Qualis classifica os veículos de publicação em oito estratos: *A1* e *A2*, estratos superiores, *B1* até *B5*, e o estrato *C* sem peso (por exemplo, novas publicações que não possuem histórico). Ainda que seja útil, é importante perceber que a classificação do Qualis é apenas para avaliar programas de pós-graduação, não sendo adequado para avaliar pesquisadores individualmente (como é o foco do trabalho).

Dado que o Fator de Impacto e o Qualis não são adequados para avaliar pesquisadores de Ciência da Computação, nesse trabalho é proposto um método para classificar os veículos de publicação e, portanto, estimar o potencial impacto de cada artigo⁹. Em suma, definida a métrica de qualidade, o próximo passo é responder as seguintes questões:

Q4: Como é a distribuição das publicações dos pesquisadores do CNPq de acordo com o método de classificação?

Q5: Como é a distribuição de pesquisadores por subáreas de Ciência da Computação considerando o método de classificação de veículos de publicação proposto?

Em seguida, a Seção 3.3.1 introduz o método de classificação de veículos de publicação, enquanto as Seções 3.3.2 e 3.3.3 respondem as duas questões mencionadas.

3.3.1 Classificação de Veículos de Publicação

Nessa seção é proposto um método de classificação de veículos de publicação baseado no índice h . Em síntese, o índice h [Hirsch, 2005] de cada veículo de publicação na base de dados (como descrito na Seção 3.1) é calculado, os veículos são ordenados e, em seguida, divididos em cinco estratos de acordo com índice h . De forma similar para pesquisadores, o índice h para um veículo de publicação é igual n , se n publicações desse veículo possuem pelo menos n citações. A Figura 3.6 mostra a distribuição do número de veículos de publicação por índice h .

Note que a maioria dos veículos de publicação possuem índice h entre 1 e 40, e apenas alguns veículos possuem valores de índice h maior que 50, formando uma cauda longa. Nesse sentido, uma abordagem natural seria dividir os veículos diretamente em faixa de índice h . Entretanto, essa abordagem produziria estratos desbalanceados,

⁹É apenas uma estimativa, pois certamente não é a única maneira de definir a qualidade de uma publicação, mas é suficiente para classificar o conjunto de publicações.

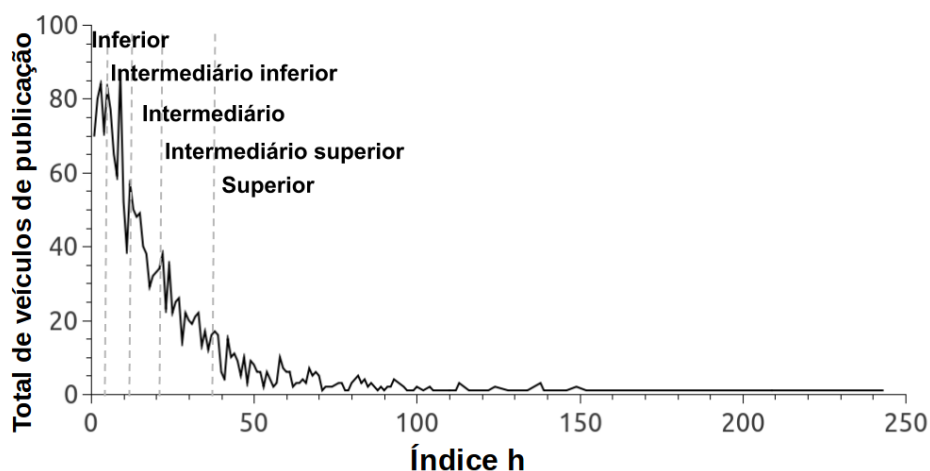


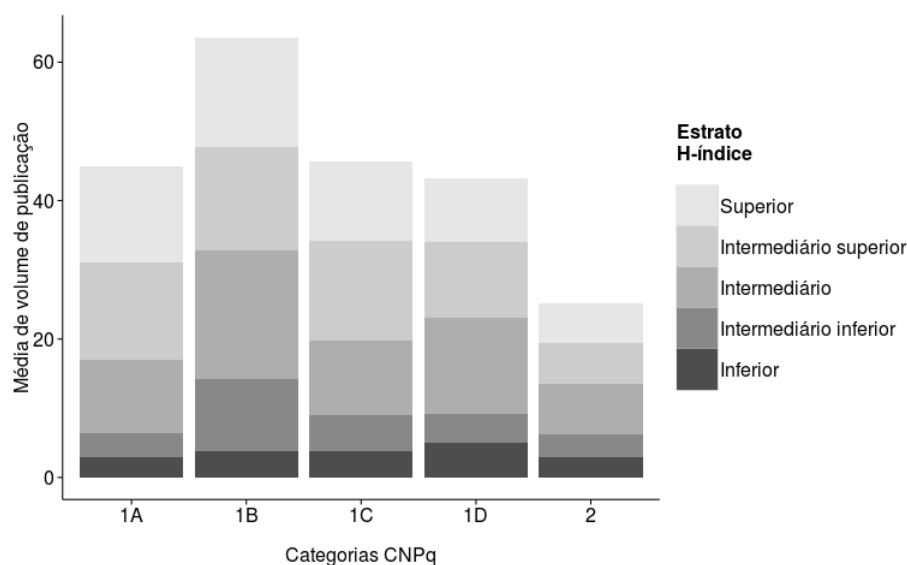
Figura 3.6. Número de veículos de publicação por índice h , com o intervalo dos estratos definidos.

como muitos veículos de publicação em estrato inferiores e apenas alguns veículos em estratos superiores. Assim, os veículos de publicação foram divididos em cinco percentis homogêneos de 20% em ordem ascendente de índice h : superior, intermediário superior, intermediário, intermediário inferior e inferior, conforme ilustrado na Figura 3.6. Em outras palavras, o estrato superior contém 20% dos veículos com maiores valores de índice h , enquanto que o estrato inferior contém os 20% dos veículos com menores valores de índice h .

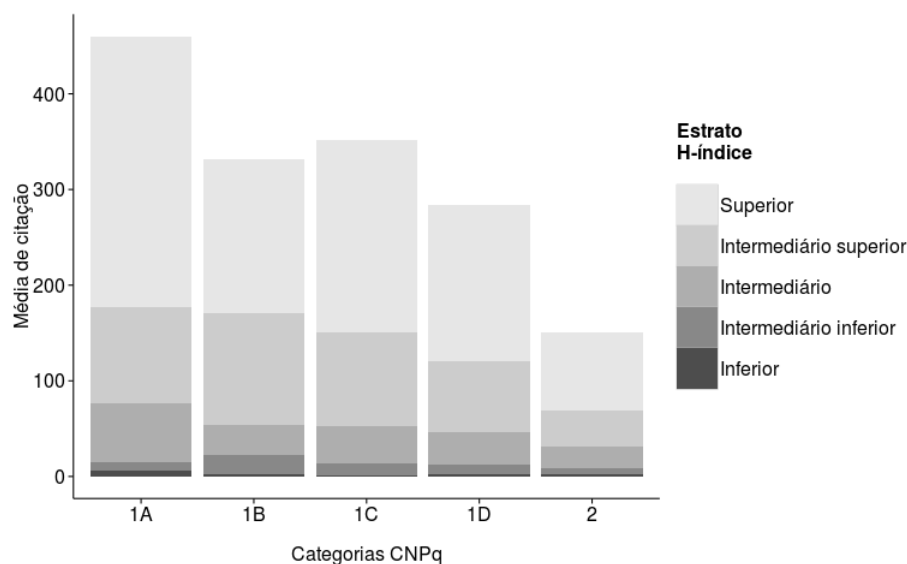
3.3.2 Qualidade das publicações com base nos estratos de índice h

Para qualificar o desempenho dos pesquisadores e responder a questão Q4, nesta seção as publicações são analisadas de acordo com os estratos de índice h ao longo das categorias do CNPq. A Figura 3.7 mostra a média de volume de publicação e quantidade de citações nos estratos de índice h em cada categoria do CNPq. A média é definida como o total de publicações na Figura 3.7 (a) e citações na Figura 3.7 (b), considerando cada estrato de índice h separadamente normalizado pelo número de pesquisadores da categoria correspondente. A Figura 3.7 (a) mostra que nem sempre os pesquisadores publicam em veículos que pertencem ao estrato superior. O aspecto crescente de publicações ao longo dos estratos de índice h é notável apenas para os pesquisadores da categoria 1A, enquanto que as demais categorias possuem maior média de publicação entre os estratos intermediário e intermediário superior.

Comparando as Figuras 3.7 (a) e (b) é possível determinar a concordância entre



(a) Volume de publicação



(b) Quantidade de citações

Figura 3.7. Evolução da média por pesquisador de (a) publicação e (b) quantidade de citações nos estratos de índice h ao longo das categorias do CNPq.

volume de publicação e quantidade de citações de acordo com os estratos de índice h . Por exemplo, os pesquisadores da categoria *1B* possuem maior volume de publicação, entretanto essa categoria não possui a maior média de citação. Em relação ao estrato superior em termos de volume de publicações, os pesquisadores das categorias *1A* e *1B* apresentam quase a mesma média, contudo em termos de quantidade de citações, pesquisadores da categoria *1A* superam os da categoria *1B* notavelmente. Analisando o

estrato superior em termos de quantidade de citações, a categoria *1C* supera a categoria *1B*, enquanto que a categoria *1D* possui a média próxima da categoria *1B*. Em termos de volume, tanto a categoria *1C* quanto a categoria *1D* apresentam taxas de publicação menores do que da categoria *1B*. Por fim, a diferença entre pesquisadores da categoria *2* para as demais categorias torna-se considerável em termos de média de volume e citação a medida que a qualidade do veículo de publicação aumenta.

Em geral, em termos de quantidade de citações, os veículos de publicação pertencentes ao estrato inferior alcançam baixo número de citações – possivelmente devido a baixa visibilidade, popularidade ou até mesmo devido a periodicidade das edições dos veículos. O baixo número de citações afeta principalmente pesquisadores da categoria *1D*, que possuem a maior media nesse estrato.

Apesar dessa análise apresentar aspectos de qualidade das publicações, não é possível afirmar que pesquisadores de uma determinada categoria possuam um conjunto de publicações melhor do que outro com base apenas nos estratos de índice h propostos. Visto que os resultados não consideram questões como: veículo de publicação que tem potencial para crescer em termos de quantidade de citações (e.g., periódicos recentemente lançados), veículos especializados em subáreas que começaram a ser exploradas recentemente (por exemplo, Nanotecnologia), veículos mantidos por comunidades pequenas (por exemplo, Computação musical). Essas questões relacionadas ao perfil de cada subárea serão discutidas na próxima seção.

3.3.3 Qualificação de publicação por subáreas

De acordo com Glänzel & Schubert [2003], Oliveira et al. [2012] e Wainer et al. [2013], o desempenho de um pesquisador pode variar significativamente dependendo de sua subárea de interesse. Por exemplo, a avaliação experimental da subárea de Interação Humano-Computador geralmente demanda mais tempo do que outras subáreas de Ciência da Computação, a partir do momento que é necessário avaliar o desempenho do usuário [Barbosa & de Souza, 2011]. Por outro lado, outras subáreas de Ciência da Computação como Bancos de Dados e Computação Gráfica não enfrentam o mesmo problema, uma vez que a avaliação experimental dessas subáreas depende de resultados gerados por um processo automático, como avaliar uma consulta ou tempo de renderização de um vídeo. Para considerar tais peculiaridades e responder a questão de pesquisa Q5, nessa seção são caracterizadas a proporção de estrato de índice h e a distribuição das categorias do CNPq em cada subárea de Ciência da Computação.

A Figura 3.8 mostra a proporção de estrato de índice h nas 23 áreas. São calculadas todas as publicações dos pesquisadores CNPq por subárea, i.e cada barra re-

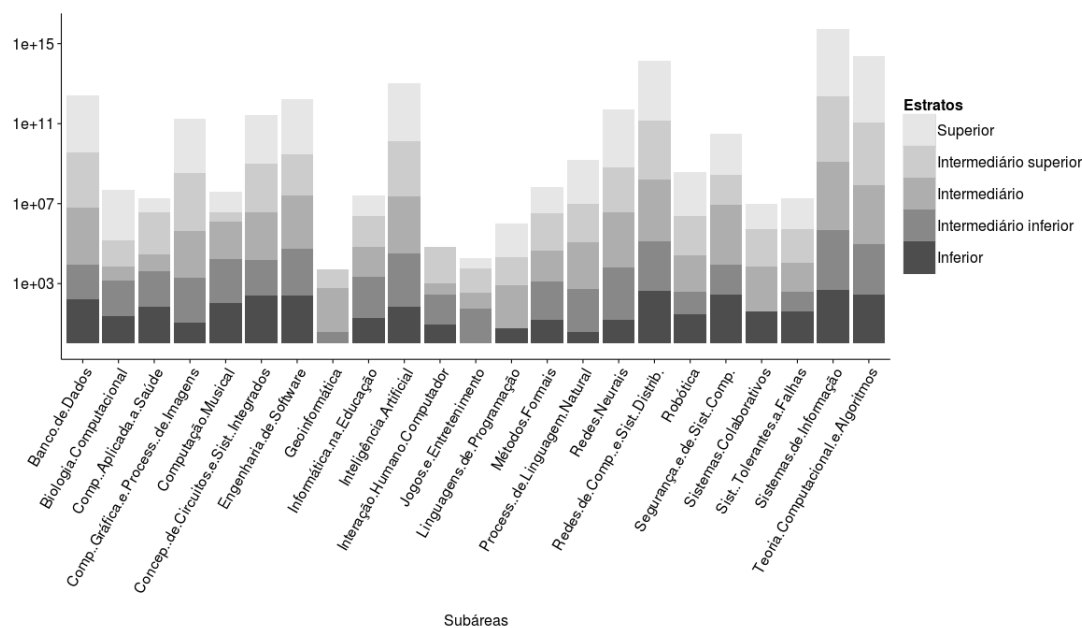


Figura 3.8. Representatividade dos estratos de índice h entre as subáreas

presenta a proporção de publicação em cada estrato, com escala logarítmica. Note que cada subárea não possui o mesmo padrão de distribuição de estrato, corroborando com os trabalhos citados no parágrafo anterior. Em geral, os pesquisadores de todas as subáreas tendem a publicar em veículos que pertencem ao estrato superior, exceto as subáreas *Jogos e Entretenimento*, *Computação Aplicada à Saúde* e *Computação Musical*, que são subáreas novas sem um amplo portfólio de veículos de publicação de impacto. Em outras subáreas a proporção de publicações cresce a medida que o nível estrato aumenta, como em *Teoria Computacional e Algoritmos* e *Sistemas de Informação*.

Não é possível afirmar que alguns pesquisadores possuem publicações melhores que outros com base apenas na distribuição de estratos, visto que cada subárea possui suas peculiaridades, como ausência de veículos de alto impacto, avaliação experimental que consome mais tempo, entre outras. Por exemplo, *Biologia Computacional* concentra suas publicações no estrato superior. Isso se deve principalmente à existência de vários periódicos de alto impacto nessa subárea (*Nature*, *Science* e *Cell*). Dessa forma, pesquisadores dessa subárea tendem a publicar menos artigos, mas com alto impacto. Por outro lado, pesquisadores da comunidade de *Geoinformática* concentram suas publicações no estrato intermediário, visto que essa subárea conta com poucos veículos de alto impacto.

A Figura 3.9 mostra a distribuição de pesquisadores CNPq em cada subárea, calculada considerando todas as publicações dos pesquisadores. Por exemplo, suponha

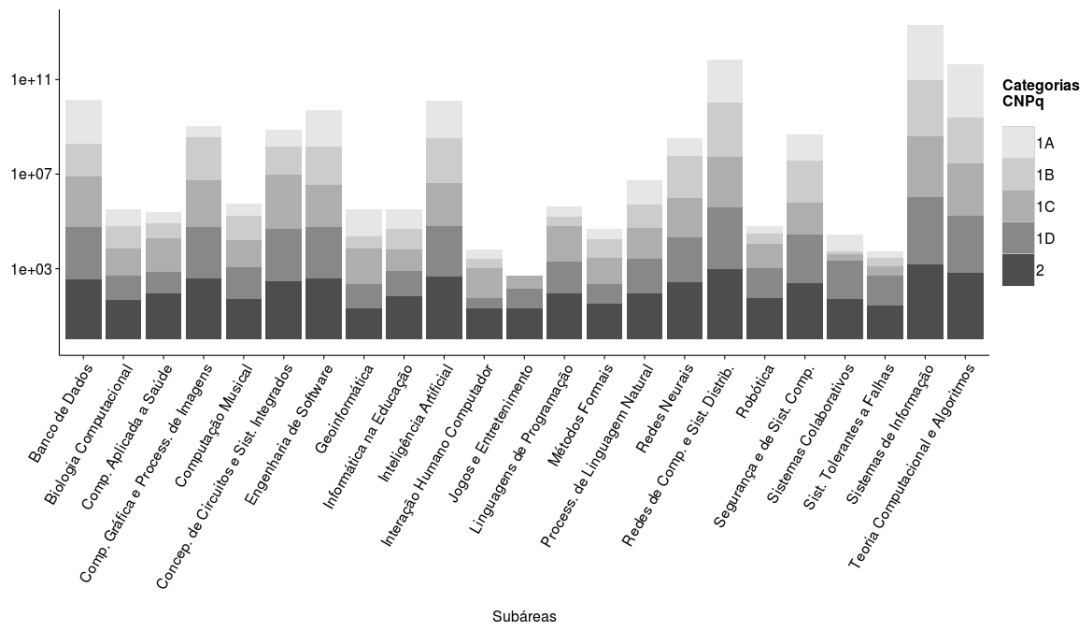


Figura 3.9. Distribuição de pesquisadores CNPq nas subáreas

que um pesquisador da categoria *1A* possua um artigo classificado em três subáreas, assim esse artigo contribui para presença dessa categoria nas três subáreas com uma pontuação de $\frac{1}{3}$. Como os valores entre as subáreas variam consideravelmente, foi utilizado escala logarítmica para gerar a Figura 3.9.

Como pode ser visto na Figura 3.9, a subárea *Jogos e Entretenimento* não possui pesquisadores das categorias *1A* e *1B*, o que pode ser explicado por essa subárea ser nova em Ciência da Computação e possivelmente não existem pesquisadores que preencham os requisitos necessários para pertencerem a categorias superiores. De forma semelhante, *Sistemas Colaborativos* e *Sistemas Tolerantes a Falhas* seguem o mesmo padrão, com a maioria dos pesquisadores da categoria *1D* e *2*.

De forma geral, de acordo com a qualidade das publicações, a maioria dos pesquisadores tendem a publicar no estrato superior, algumas exceções podem ser explicadas pelas peculiaridades de cada subárea. Algumas subáreas publicam mais do que outras, como é o caso de *Teoria Computacional* e *Sistemas de Informação*¹⁰. Por outro lado, subáreas como *Jogos e Entretenimento* e *Geoinformática* possuem menos publicações, provavelmente devido ao tamanho de suas comunidades de pesquisadores. Assim, subáreas recém exploradas em Ciência da Computação podem sofrer de pouca representação ou nenhuma de categorias superiores. Especificamente, a ausência de pesquisadores da categoria *1A* e *1B* pode indicar que essas subáreas recebem menos financiamento, o

¹⁰Note que a maior parte dos veículos de publicação são classificados nessas duas subáreas, assim, não é surpreendente que essas subáreas concentrem a maior parte da base de dados.

que pode impactar no desempenho de pesquisadores. Por outro lado, subáreas como *Algoritmos e Teoria*, *Rede de Computadores*, *Sistemas de Informação* e *Engenharia de Software* possuem pesquisadores de todas as categorias do CNPq, o que indica que são subáreas consolidadas em Ciência da Computação.

3.4 Considerações finais

Neste capítulo foi apresentada a base de dados utilizada no trabalho, bem como os passos para construí-la de forma que fosse possível avaliar os pesquisadores brasileiros de Ciência da Computação. Com o conjunto de dados definido, foram analisados os perfis de 406 pesquisadores que participam do Programa de Bolsas de Produtividade do CNPq. Em tal análise foram consideradas três dimensões: tempo de carreira (tempo desde que o pesquisador obteve título de doutor), orientação de alunos (número de alunos de mestrado e doutorado orientados) e produção científica (volume de publicação e quantidade de citações). Em relação à produção científica, foi proposto um método para inferir o impacto de veículos de publicação. Assim, a produção científica foi analisada também em termos do impacto do veículo de publicação e de acordo com o padrão de publicação de cada subárea de Ciência da Computação.

Em geral, as seguintes conclusões foram alcançadas com esse estudo:

- Em termos de tempo de carreira, quanto mais experiente é o pesquisador, mais elevada é sua posição no *ranking* do CNPq, apesar de existirem pesquisadores com mais de 25 anos de carreira na categoria 2. Além disso, normalmente demanda mais tempo para o pesquisador ser promovido da categoria 2 para categoria 1D e da categoria 1B para categoria 1A.
- A proporção de alunos de mestrado orientados durante a carreira do pesquisador é perto de um estudante por ano de carreira, proporção similar em todas as categorias do CNPq. Entretanto, a proporção de alunos de doutorado orientados durante o tempo de carreira do pesquisador se diferencia entre as categorias do CNPq. O valor da mediana de alunos de doutorado decresce do nível 1 para nível 2, a exceção ocorre para a categoria 1A que possui o valor da mediana similar à mediana da categoria 1D.
- Em relação à distribuição de volume de publicação e quantidade de citações, as categorias do CNPq seguem a distribuição esperada, ou seja, o valor da mediana decresce ao longo das categorias. Novamente, a exceção é a categoria 1A que possui o valor da mediana similar ao da categoria 1C para os dois critérios.

- Os pesquisadores avaliados nem sempre escolhem publicar em veículos de publicação que pertencem aos estratos superiores. Com exceção dos pesquisadores da categoria *1A* os quais publicam em veículos pertencentes ao estrato superior, os pesquisadores das demais categorias dedicam seus esforços em veículos dos estratos intermediário e intermediários superior.
- Apesar da maioria das subáreas de Ciência da Computação concentrarem suas publicações em veículos do estrato superior, algumas peculiaridades de cada subárea (tamanho da comunidade, ausência ou a presença de veículos de publicação de alto impacto, ausência de pesquisadores experientes, etc) afetam a distribuição de publicações em termos dos estratos de impacto propostos.

A análise apresentada referente às subáreas de Ciência da Computação também mostrou que o padrão de publicação varia de acordo com cada subárea. Como observado na Tabela 3.1, cada subárea tem perfis diferentes, algumas com dedicação em conferências outras em periódicos, além da média de volume de publicação e quantidade de citações variarem dentro de cada subárea e entre as subáreas. Como mostrado nas Seções 3.3.3 e 3.3.2, cada subárea possui particularidades que afetam o padrão de publicação. Neste sentido, no próximo capítulo são propostas premissas para aferir um viés imparcial em uma avaliação de pesquisadores de diferentes subáreas. Além disso, é apresentado o método de *ranking* de pesquisadores que implementa tais premissas.

Capítulo 4

Ranqueamento de pesquisadores considerando múltiplas subáreas de pesquisa

Avaliar o impacto da pesquisa científica é uma tarefa desafiadora, que levou a criação de vários índices de produtividade, conforme discutido no Capítulo 2. Entretanto, a maioria desses índices desconsidera a especificidade de diferentes áreas de pesquisa e avalia os pesquisadores com base em indicadores globais, como número de publicações e citações. Tais indicadores podem variar consideravelmente entre áreas, ou até mesmo entre subáreas de pesquisa, se tornando inviáveis para avaliar perfis de áreas distintas. Dessa forma, com objetivo de propiciar uma avaliação mais justa da produção científica de pesquisadores, neste trabalho é proposta uma abordagem de avaliação que considera todas as subáreas de atuação de um pesquisador e pondera as características de cada subárea. A Seção 4.1 apresenta as premissas consideradas para assegurar uma avaliação imparcial de pesquisadores de subáreas distintas, e a Seção 4.2 apresenta o método de *ranking* proposto.

4.1 Premissas de avaliação de produtividade entre subáreas distintas

Conforme apresentado na avaliação dos pesquisadores brasileiros de Ciência da Computação no Capítulo 3 e nos demais estudos de avaliação de produtividade no Capítulo 2, o perfil de publicação de cada subárea está ligado a diversos fatores como portfólio de veículos de publicação, tamanho da comunidade, distribuição de pesquisadores de

diversos níveis de experiência, entre outros. Nesse sentido, a diversidade de cada subárea influencia diretamente no padrão de publicação do pesquisador. Para aumentar a complexidade de se avaliar adequadamente a produtividade científica, um pesquisador ao longo de sua carreira acadêmica pode ter publicações ligadas a diversas subáreas dentro de uma área científica, aferindo um perfil plural ao conjunto de publicações do pesquisador [Wagner et al., 2011]. Assim, em uma avaliação de pesquisadores mais justa deve considerar tanto a pluralidade do pesquisador quanto a diversidade de cada subárea, sem desmerecer uma subárea em relação às outras.

A fim de propiciar uma avaliação imparcial da produção científica entre pesquisadores de subáreas distintas, um índice de produtividade deve seguir propriedades que aferem um viés justo e imparcial em uma avaliação, respeitando as idiosincrasias de cada subárea. Nesse sentido, são propostas três premissas norteadoras do método de *ranking* desenvolvido nesta dissertação:

- *Pluralidade.* A produtividade de um pesquisador deve ser avaliada em todas as subáreas em que o pesquisador tenha publicado. Ao longo de sua carreira, é comum que um pesquisador atue em mais de uma subárea. Além disso, algumas subáreas de pesquisa possuem alto grau de multidisciplinaridade, fazendo com que o pesquisador publique seus trabalhos em diversas comunidades. Por exemplo, é comum que um pesquisador cuja subárea de atuação principal seja Robótica também possua trabalhos relacionados à área de Inteligência Artificial, pois várias soluções para problemas de Robótica provêm de Inteligência Artificial. Assim, em uma avaliação justa de pesquisadores de subáreas distintas, devem ser consideradas todas (ou a maioria) as subáreas de atuação do pesquisador e ponderar suas respectivas particularidades, conforme a próxima premissa.
- *Diversidade.* O perfil de cada subárea de pesquisa deve ser considerado em uma avaliação de produtividade do pesquisador. Como mostrado nas análises apresentadas no Capítulo 3, cada subárea possui um padrão de publicações, como número de artigos, quantidade de citações, concentração em veículos de diferentes níveis de qualidade, etc. Dessa forma, a comparação direta de índices de produtividade entre subáreas distintas pode ser desbalanceada e privilegiar indevidamente algumas subáreas em detrimento a outras. Portanto, um método de *ranking* com pesquisadores de diferentes subáreas deve considerar adequadamente tais peculiaridades.
- *Igualdade.* As subáreas de pesquisa devem ser consideradas igualmente importantes e merecedoras de mérito científico. Como mencionado anteriormente, cada

subárea possui seu padrão de publicação considerando: tamanho da comunidade científica, complexidade da metodologia de avaliação, dedicação em periódicos ou conferência, etc. Assim, subáreas que tenham média de citações maior do que outras não devem ser consideradas de maior prestígio, pois tal número pode ser em consequência do tamanho da comunidade (em que uma comunidade maior possui mais autores e mais publicações que possam citar os trabalhos). De forma semelhante, subáreas que produzam mais artigos não são necessariamente mais eficientes do que outras com menor número de artigos (novamente pode ser em relação ao tamanho da comunidade, ao número de veículos de publicação, etc). Diante disso, um método de avaliação deve considerar que as subáreas são igualmente importantes, independente de possíveis discrepâncias em tais valores absolutos.

Especificamente no método proposto, a premissa *pluralidade* é considerada mapeando as posições dos pesquisadores nos *rankings* produzidos separadamente para cada subárea. Em seguida, as posições dos pesquisadores em cada *ranking* são comparadas utilizando percentil. No contexto de *ranking*, percentil é o percentual de itens que são iguais ou menores que um determinado item no *ranking*. Dessa forma, o desempenho do pesquisador em cada subárea não é confrontado diretamente, evitando comparações indevidas de índices de produtividade de subáreas distintas. Para assegurar a *igualdade* entre as subáreas, é proposta uma nova abordagem de agregação de *ranking* que garanta o tratamento igualitário das subáreas. A próxima seção descreve detalhadamente como tais premissas são implementadas no método proposto.

4.2 Método de ranqueamento de pesquisadores entre subáreas distintas

Para superar as limitações dos índices de produtividade existentes e atender as premissas descritas na seção anterior, é proposta uma nova abordagem que avalia pesquisadores em diversas subáreas de pesquisa, agregando evidências de produtividade de cada subárea separadamente. Inspirado em abordagens de agregação de *ranking* em recuperação da informação distribuída [Callan, 2000], a abordagem proposta é genérica e pode ser utilizada para produzir versões de índices globais em índices comparáveis entre subáreas (e.g, volume de publicação, número de citação, e até mesmo o índice *h*).

Em particular, a abordagem proposta visa produzir um *ranking* global de pesquisadores em uma dada área científica, mapeando a posição de cada pesquisador nos

rankings construídos para diversas subáreas dentro da mesma áreas científica. Desse modo, a *pluralidade* do padrão de publicação do pesquisador é considerada. Para formalizar a abordagem proposta, considere p_i^s como a pontuação atribuída ao pesquisador i em uma determinada subárea s . Essa pontuação é calculada somando a contribuição de cada publicação do pesquisador para a subárea s , considerando que cada publicação possa ser classificada em várias subáreas, com base, por exemplo, nas subáreas de interesse do veículo de publicação em que o artigo foi publicado. Diante disso, a pontuação atribuída a cada pesquisador é definida pela Equação 4.1.

$$p_i^s = \sum_{j=1}^{n_i} \mathbf{1}_j^s \frac{p_{i,j}}{q_{i,j}}, \quad (4.1)$$

onde n_i é o número total de publicações do pesquisador i , $\mathbf{1}_j^s$ é uma função indicadora (assume valor 1 se a j -ésima publicação do pesquisador pertence à subárea s , e 0 caso contrário), $p_{i,j}$ é a pontuação dada pela j -ésima publicação ao pesquisador i ¹, e $q_{i,j}$ é a quantidade total de subáreas em que a publicação j é classificada. O fator $q_{i,j}$ age como normalizador, a fim de garantir que a pontuação $p_{i,j}$ não seja considerada mais de uma vez quando a pontuação do pesquisador for agregada entre as diversas subáreas.

Pesq.	Citações	Subárea
r_1	18	BD GEO SI
	10	BD SI
	2	BD GEO

Pesq.	Citações	Subárea
r_2	10	GEO SI
	4	BD GEO
	2	BD GEO
	14	SI BD

Pesq.	Citações	Subárea
r_3	10	GEO SI
	2	GEO
	6	SI BD GEO
	7	BD

Tabela 4.1. Conjunto de citações dos pesquisadores r_1 , r_2 e r_3 .

Para elucidar o método proposto, considere os conjuntos de publicações dos pesquisadores r_1 , r_2 e r_3 apresentados na Tabela 4.1. Note que os três autores possuem publicações em três subáreas: Bancos de Dados (BD), Geoinformática (GEO) e Sistemas de Informação (SI). De forma geral, essas publicações contribuem para os pesquisadores no *ranking* dessas subáreas. Considerando a quantidade de citações como critério de *ranking* e de acordo com a Equação 4.1, as pontuações atribuídas aos pesquisadores são:

¹Para a pontuação baseada em volume, tem-se $p_{i,j} = 1, \forall j \in [1, n_i]$; para pontuação baseada em citação, $p_{i,j}$ é a quantidade total de citações da j -ésima publicação do pesquisador i .

$$\begin{array}{ccc}
 p_{r_1}^{BD} = \frac{18}{3} + \frac{10}{2} + \frac{2}{2} = 12 & \left| & p_{r_2}^{BD} = \frac{4}{2} + \frac{2}{2} + \frac{14}{2} = 10 \\
 p_{r_1}^{GEO} = \frac{18}{3} + \frac{2}{2} = 7 & \left| & p_{r_2}^{GEO} = \frac{10}{2} + \frac{4}{2} + \frac{2}{2} = 8 \\
 p_{r_1}^{SI} = \frac{18}{3} + \frac{10}{2} = 11 & \left| & p_{r_2}^{SI} = \frac{10}{2} + \frac{14}{2} = 12 \\
 \text{Pesquisador } r_1 & & \text{Pesquisador } r_2
 \end{array}
 \quad \left| \quad \begin{array}{ccc}
 p_{r_3}^{BD} = \frac{6}{3} + \frac{7}{1} = 9 & & \\
 p_{r_3}^{GEO} = \frac{10}{2} + \frac{2}{1} + \frac{6}{3} = 9 & & \\
 p_{r_3}^{SI} = \frac{10}{2} + \frac{6}{3} = 7 & & \\
 \text{Pesquisador } r_3 & &
 \end{array}$$

Dadas as diferenças entre as subáreas, as pontuações produzidas para os pesquisadores podem não ser comparáveis. Com objetivo de garantir que a *diversidade* das subáreas seja respeitada, foi utilizado o percentil do pesquisador em cada área, em vez de se utilizar a pontuação absoluta atribuída ao pesquisador. Como resultado, foi considerada a posição do pesquisador relativa aos outros pesquisadores da subárea. Para formalizar a intuição, o percentil $percentil_i^s$ do pesquisador i na subárea s é definido pela Equação 4.2.

$$percentil_i^s = \frac{m_i^s + 0.5e_i^s}{N^s}, \quad (4.2)$$

onde N^s é o total de pesquisadores na subárea s , m_i^s e e_i^s são as quantidades de pesquisadores com pontuação menor ou igual ao pesquisador i , respectivamente. O último fator garante que pesquisadores com pontuações iguais sejam considerados adequadamente [Hyndman & Fan, 1996].

Por fim, é necessário garantir que a *igualdade* entre as subáreas seja mantida. Para tanto, a abordagem equipara o percentil de um pesquisador em uma subárea específica com o mesmo percentil nas demais subáreas em que o pesquisador atua. Formalmente, considere b_i como a subárea base do pesquisador i , i.e., subárea em que o pesquisador possui o maior valor de percentil $percentil_i^s$ de acordo com a Equação 4.3.

$$sb_i = \arg \max_{s \in S} percentil_i^s, \quad (4.3)$$

onde S representa o conjunto de todas as subáreas que estão sendo consideradas na área científica em questão.

Para agregar as pontuações obtidas pelos pesquisadores nas múltiplas subáreas, o percentil obtido pelo pesquisador em cada subárea é projetado em sua subárea base. Como resultado, o pesquisador é recompensado com a pontuação da subárea base correspondente ao percentil obtido nas demais subáreas. Por exemplo, um pesquisador com percentil 0,90 em Banco de Dados e 0,30 em Computação Gráfica deve ser recompensado com a pontuação correspondente ao percentil 0,90 e 0,30 em Banco de Dados, assumindo essa seja sua subárea base. Formalmente, a pontuação agregada p_i entre as

subáreas de um pesquisador i de acordo com o conjunto de subáreas $s \in S$ em uma área científica de interesse é definida pela Equação 4.4.

$$p_i = \sum_{s \in S} \frac{f_{sb}(\text{percentil}_i^s)}{f_{sb}(1.0)}, \quad (4.4)$$

onde a função de projeção $f_{sb}(\text{percentil}_i^s)$ mapeia o percentil percentil_i^s obtido pelo pesquisador i na subárea s para a pontuação correspondente a esse percentil na subárea base sb do pesquisador, e $f_{sb}(1.0)$ retorna a pontuação máxima em sb entre todos os pesquisadores nessa subárea. Essa normalização elimina qualquer viés de pesquisadores com subárea base mais prolífica (e.g., subáreas com alto volume de publicação e citações), compelindo a igualdade entre subáreas.

Subárea	Pesq.	Pont.	Percentil(%)	Subárea	Pesq.	Pont.	Percentil(%)
BD	r_1	12	67	GEO	r_3	9	67
	r_2	10	33		r_2	8	33
	r_3	9	0		r_1	7	0

Subárea	Pesq.	Pont.	Percentil(%)
SI	r_2	12	67
	r_1	11	33
	r_3	7	0

Tabela 4.2. *Rankings* produzido para cada subárea de acordo com o desempenho dos pesquisadores r_1 , r_2 e r_3 .

Ainda de acordo com os conjuntos de publicações da Tabela 4.1 e as pontuações atribuídas aos pesquisadores em suas três subáreas de atuação, são definidos os *rankings* das três subáreas conforme apresentados na Tabela 4.2. Os valores da coluna Percentil são definidos pela Equação 4.2 e de acordo com desempenho dos três pesquisadores em questão.

As subáreas base dos pesquisadores r_1 , r_2 e r_3 de acordo com a Equação 4.3 são Bancos de Dados, Sistemas de Informação e Geoinformática, respectivamente. Em seguida, é necessário agregar as pontuações obtidas pelos pesquisadores em suas subáreas de atuação. Por exemplo, a pontuação global do pesquisador r_1 é obtida somando o valor 12, correspondente ao valor obtido pelo pesquisador em sua subárea base Bancos de Dados, com as pontuações do *ranking* de BD correspondentes aos percentis do pesquisador em suas demais subáreas de atuação. Portanto, devem ser somados à pontuação global do pesquisador r_1 os valores 9 e 10, correspondentes aos percentis 0 e 0,33 de BD, referente ao desempenho do pesquisador em Geoinformática e Sistemas de Informação, respectivamente. Por fim, o valor total deve ser normalizado pela maior pontuação

alcançada por um pesquisador da subárea base de r_1 . Dessa forma, a pontuação global do pesquisador r_1 é dada por $\frac{12+9+10}{12} = 2,58$. Repetindo o mesmo processo para os demais pesquisadores, a pontuação global de r_2 e r_3 é 2,83 e 2,55, respectivamente. Assim, o método proposto ordena os pesquisadores da seguinte forma: r_2 , r_1 e r_3 .

Capítulo 5

Avaliação experimental

Este capítulo apresenta a metodologia experimental empregada para avaliar o método de *ranking* proposto e os resultados obtidos de tal avaliação. Primeiramente, é necessário avaliar se o método proposto realmente provê melhorias frente aos índices comumente utilizados em tarefas de avaliações de pesquisadores. Em seguida, é pertinente identificar quais outros fatores podem influenciar na abordagem proposta e como seria o comportamento do método diante da situação real de promoção de pesquisadores.

Dessa forma, o método primeiramente é confrontado com os índices globais, como o índice h , quantidade de citações e volume de publicação. Essa análise verifica se o *ranking* gerado com a abordagem proposta é de fato melhor do que o *ranking* produzido com índices globais. Apesar do bom resultado obtido com essa análise, os resultados mostraram que os índices globais conseguem melhores desempenhos nas primeiras posições dos *rankings*, motivando a combinação de índices globais com a abordagem proposta. Com essa combinação é possível explorar evidências globais não contempladas pela abordagem e que, provavelmente, ajudem a gerar um *ranking* mais próximo do *ranking* de referência, principalmente nas posições iniciais.

Em seguida, ainda que a qualidade do *ranking* gerado com a abordagem tenha sido satisfatória, outros fatores normalmente utilizados em avaliações de pesquisadores (orientação de alunos, liderança de grupos de pesquisa, inserção internacional, etc) podem influenciar nos resultados do método. Assim, os casos de falha do método são analisados afim de identificar tais fatores. Além da análise dos casos de falha, o comportamento do método ainda é avaliado frente às promoções recebidas por pesquisadores.

Em resumo, são propostas quatro questões de pesquisas, que guiam a avaliação experimental do método proposto para *ranking* de pesquisadores:

Q1. É possível melhorar os índices de produtividade existentes com a abordagem

proposta que considera o perfil distinto de subáreas? Para tanto, os *rankings* gerados pelo método proposto são comparados com os *rankings* gerados com índices globais.

- Q2. É possível combinar de forma efetiva índices globais com a abordagem proposta? Essa questão é abordada combinando linearmente o *ranking* gerado pela abordagem proposta com o *ranking* produzido por índices globais.
- Q3. Quais outros fatores podem afetar a abordagem proposta? Para responder essa questão é feita uma análise dos casos em que o método proposto falha, identificando os pontos que de certa forma interferem nos resultados obtidos.
- Q4. É possível identificar pesquisadores a serem promovidos com os resultados obtidos pela abordagem? Para responder essa pergunta são analisados os pesquisadores que foram promovidos de categoria na avaliação mais recente do CNPq.

Esse capítulo está organizado da seguinte maneira. A seção 5.1 apresenta a base de dados e as características que possibilitam sua utilização na avaliação experimental. A Seção 5.2 revisa as métricas de produtividades que são comparadas com o método proposto. Em seguida, a Seção 5.3 define o *ranking* de referência e a métrica de avaliação empregada para reportar os resultados. Por fim, a Seção 5.4 mostra os resultados obtidos, respondendo as questões de pesquisas propostas.

5.1 Conjunto de publicações

Índices bibliométricos normalmente dependem de uma base de dados de publicações que contenha informações de qualidade e confiáveis acerca das publicações. Bibliotecas digitais, como DBLP e *IEEE Xplore*¹ fornecem informações sobre autores, títulos, veículos de publicação e ano de publicação. Entretanto, tais plataformas não fornecem a quantidade de citações de cada publicação ou o índice *h* dos pesquisadores. Além disso, máquinas de busca voltadas para pesquisa científica (como *Google Scholar*², *ArnetMiner* e *Microsoft Academic Search*) fornecem o índice *h* e outros índices bibliométricos, mas a organização dessas plataformas referente à autoria das publicações pode conter imprecisões. Por exemplo, o agrupamento das publicações de cada pesquisador pode estar errado devido à ambiguidade de nomes: o mesmo autor pode ser referenciado com nomes distintos ou diferentes autores podem apresentar nomes similares. Além disso,

¹IEEE Xplore: <http://ieeexplore.ieee.org/>

²Google Scholar: <http://scholar.google.com.br/>

a coleção indexada por essas máquinas de busca reúne todos documentos de autoria do pesquisador, incluindo trabalhos não científicos, relatórios técnicos e até mesmo reportagens de revistas não científicas.

Desse modo, a base de dados apresentada no Capítulo 3 é utilizada para avaliar o método proposto por conter as seguintes características: inclui apenas publicações que apareceram em conferências e periódicos; agrupa corretamente publicações e seus autores; fornece a quantidade de citações de cada publicação (permitindo o cálculo do índice h de cada pesquisador) e classifica cada publicação em uma ou várias subáreas.

5.2 Rankings de comparação

Dado o conjunto de publicações, é necessário definir os índices (ou métricas) de produtividades que serão utilizados para comparar com método proposto. Foram escolhidos os mesmos índices utilizados em plataformas acadêmicas (como *ArnetMiner*, *Microsoft Academic Search* e *Google Scholar*) e em diversos trabalhos de avaliação de produtividade [Helene & Ribeiro, 2011; Kato & Ando, 2013; Almeida & Guimarães, 2013; Vanecek, 2014], tais índices são: volume de publicação, número de citações e o índice h [Hirsch, 2005]. Em volume de publicação, os pesquisadores são ordenados em ordem decrescente do número total de publicações. De forma similar, em número de citações, os pesquisadores são ordenados em ordem decrescente em relação à quantidade total de citações recebidas por seu conjunto publicações. Por fim, em relação ao índice h , os pesquisadores são ordenados de forma decrescente em relação ao seu valor de índice h .

5.3 Procedimento de avaliação

Para mensurar o método proposto frente aos índices globais, é necessário definir um *ranking* de referência e uma métrica de avaliação para reportar os experimentos. O *ranking* de referência deve ser considerado ideal, visto que foi elaborado por um processo metódico realizado por um conjunto de especialistas. Assim, os *rankings* produzidos tanto com abordagem proposta quanto com os índices globais são comparados com o *ranking* de referência produzido pelo CNPq para a distribuição das Bolsas de Produtividade em Pesquisa. Por sua vez, a métrica de avaliação deve caracterizar adequadamente o *ranking* de referência. A métrica de avaliação utilizada nos experimentos considera itens de relevâncias não binárias, adequada para o mapeamento das cinco categorias de *ranking* do CNPq.

A Seção 5.3.1 define o *ranking* de referência utilizado, enquanto que a Seção 5.3.2 descreve a métrica de avaliação utilizada para reportar os resultados experimentais.

5.3.1 Ranking de referência

O primeiro passo para a configuração experimental é definir um *ranking* de referência para inferir a qualidade do *ranking* gerado com o método proposto. Como não existe um *ranking* de pesquisadores com abrangência mundial, foi decidido utilizar uma avaliação apenas com pesquisadores brasileiros. Dessa forma, para avaliar o método proposto é utilizado o *ranking* produzido pelo CNPq. Conforme descrito no Capítulo 3, a avaliação do CNPq é feita por comissões especiais de cada área de pesquisa, que divide os pesquisadores em duas categorias: categoria 1 que é dividida em 4 subcategorias (1A, 1B, 1C e 1D), em ordem decrescente de valor de financiamento, e categoria 2. Assim, um método que gere um *ranking* ideal deve colocar os pesquisadores da categoria 1A nas primeiras posições, seguidos pelos pesquisadores da categoria 1B e assim por diante. Note que dessa forma pesquisadores de mesma categoria são igualmente relevantes. Ou seja, a qualidade do *ranking* não é alterada pela troca de posições de pesquisadores de mesma categoria.

Embora o CNPq considere não apenas o conjunto de publicações em sua avaliação (e.g., proposta de projeto de pesquisa e inserção internacional), o comitê de Ciência da Computação responsável pela avaliação enfatiza que um conjunto qualificado de publicações é requisito fundamental para promover um pesquisador de categoria³. Além disso, dado que o *ranking* é produzido por comitês especializados, é pertinente utilizá-lo como *ranking* de referência na avaliação da abordagem proposta. Assim, foram consideradas as publicações dos pesquisadores classificados pelo CNPq nas categorias 1 e 2 em 2012.

5.3.2 Métrica de avaliação

Para comparar o método de *ranking* proposto aos índices de produtividades descritos na Seção 5.2, é utilizada a métrica *discounted cumulative gain* (DCG) [Järvelin & Kekäläinen, 2002]. Essa métrica tem sido amplamente utilizada para avaliar abordagem de *rankings* em várias tarefas relacionadas à recuperação de informação (Clarke et al. [2009]). Em particular, a métrica DCG adota a premissa de relevância não binária, avaliando um *ranking* com base em uma escala de relevância, de itens mais relevantes aos menos relevantes. Além disso, essa métrica aplica um fator logaritmo para que itens

³Crerios de julgamento do Comitê de Assessoramento da Ciência da Computação: <http://www.cnpq.br/web/guest/criterios-de-julgamento>

relevantes das primeiras posições do *ranking* tenham maior peso em relação aos itens menos relevantes das posições inferiores. Especificamente para análise experimental, essa métrica privilegia o *ranking* que contiver os pesquisadores da categoria *1A* nas primeiras posições, seguidos pelos pesquisadores *1B*, e assim por diante até chegar aos pesquisadores da categoria *2*, que têm menor peso. Formalmente, a métrica DCG para uma posição k do *ranking* é definida pela Equação 5.1.

$$\text{DCG}@k = \sum_{i=1}^k \frac{2^{g_i} - 1}{\log_2(i + 1)}, \quad (5.1)$$

onde g_i denota a relevância não binária associada ao item na posição i .

Como mencionado, a escala de relevância é baseada na classificação realizada pelo CNPq referente aos pesquisadores, como definida na Seção 5.3.1. É considerado o mapeamento das categorias do CNPq *1A*, *1B*, *1C*, *1D* e *2* para classes de relevância 5, 4, 3, 2 e 1, respectivamente. Por fim, para manter o valor da métrica DCG no intervalo de $[0,1]$, foi utilizada a versão normalizada da métrica DCG, denominada nDCG. Assim, para obter os valores de nDCG é necessário dividir o valor de $\text{DCG}@k$ definido na Equação 5.1 pelo melhor valor na mesma posição de corte k . Nos experimentos realizados, os valores de nDCG são reportados com vários valores de corte k , em um intervalo de 5 a 100, variando de 5 em 5.

5.4 Resultados

Seguindo as questões propostas no início deste capítulo, a Seção 5.4.1 responde a questão de pesquisa *Q1*, aplicando o método proposto em índices globais para produzir versões de índices de produtividade comparáveis entre pesquisadores de subáreas distintas. A Seção 5.4.2 discute a questão *Q2*, analisando o impacto do volume de publicação e citações como evidências para produzir *ranking* de pesquisadores. A Seção 5.4.3 responde a questão de pesquisa *Q3*, realizando uma análise do método proposto em relação ao *ranking* produzido pelo CNPq. Por fim, a Seção 5.4.4 avalia o método de acordo com as promoções dadas aos pesquisadores em uma avaliação mais recente promovida pelo CNPq, respondendo a questão de pesquisa *Q4*.

5.4.1 Ranking utilizando índice único

Para responder a questão de pesquisa *Q1* acerca da qualidade do *ranking* produzido com o método proposto, a abordagem é comparada utilizando o índice de produtividade

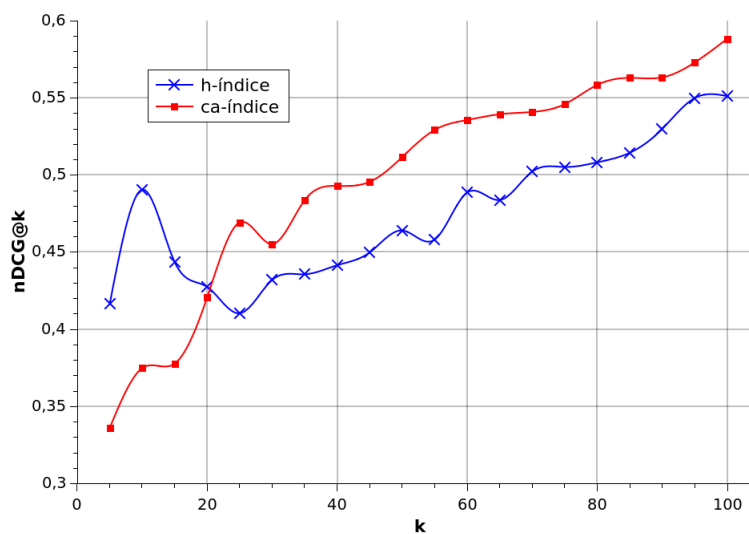


Figura 5.1. Comparação entre o *ranking* produzido com índice h e com ca -índice.

h [Hirsch, 2005]. Nesse caso, a pontuação é baseada no índice h , ou seja, $p_{i,j}$ da equação 4.1, definida no Capítulo 4, é o índice h do pesquisador. Assim, para obter a versão do índice h comparável entre pesquisadores de subáreas distintas, denominado nesta seção de ca -índice⁴, o índice h do pesquisador é propagado para cada uma de suas publicações, e de cada publicação para as subáreas nas quais a publicação é classificada, de acordo com a Equação 4.4.

A Figura 5.1 mostra a comparação entre o *ranking* produzido pelo ca -índice em oposição ao produzido com a versão original do índice h , em termos de valores de $nDCG@k$ atribuído a cada *ranking*, para valores de corte k variando de 5 a 100. É possível observar que o ca -índice supera o índice h para todo o intervalo de k . A exceção ocorre para $k \leq 20$, em que o índice h tradicional obteve melhor desempenho. Após uma análise mais detalhada acerca desse resultado, a exceção observada no topo do *ranking* se deve à presença de pesquisadores relevantes com alto nível de especialização em apenas algumas subáreas de pesquisa, ou seja, pesquisadores da categoria 1A com poucas subáreas de atuação. De fato, no *ranking* do CNPq vários pesquisadores da categoria 1A têm como foco *apenas* a subárea de Teoria Computacional. Dessa forma, a normalização por subáreas realizada pelo método proposto não consegue bons resultados para posições iniciais do *ranking*.

A Figura 5.2 mostra a distribuição dos valores do índice h antes e depois de se aplicar o método proposto: cada ponto é um pesquisador agrupado em uma das cinco categorias do CNPq, e a barra horizontal é o valor da mediana. Afim de manter os va-

⁴ ca -índice em tradução do termo utilizado nos artigos derivados desta dissertação que estão em inglês: ca -index para *cross-area index*

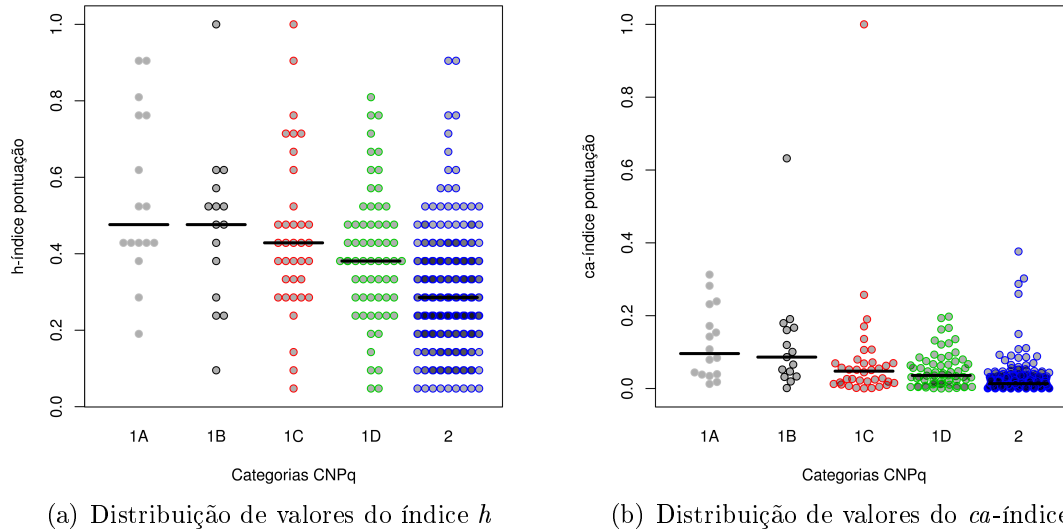


Figura 5.2. Distribuição de valores do (a) índice h e (b) ca -índice entre os cinco níveis de relevância considerados.

lores entre $[0,1]$, a pontuação de cada pesquisador é normalizada pela maior pontuação obtida por um pesquisador. A Figura 5.2(a) mostra a distribuição dos valores do índice h dos pesquisadores antes de aplicar o método proposto, enquanto a Figura 5.2(b) mostra depois de aplicado o método. A normalização entre subáreas realizada pelo método proposto reduz a dispersão de valores entre os pesquisadores classificados pelo CNPq. Tal normalização pode ter suavizado o desempenho dos pesquisadores individualmente; ou seja, é possível que algum pesquisador com alto nível de especialização e com destacado conjunto de publicações tenha sido prejudicado. Por outro lado, o método proporciona uma caracterização mais clara dos pesquisadores em cada nível, e uma melhor diferenciação entre pesquisadores de categorias distintas, como indicado estritamente pelo valor descendente da mediana a partir do nível $1A$ ao nível 2 na Figura 5.2(b).

Para investigar os resultados obtidos com índice h mostrados nas Figuras 5.1 e 5.2, o impacto da abordagem proposta é avaliado utilizando separadamente as evidências bases do índice h : volume de publicação e citação. A Figura 5.3 apresenta o nDCG dos *rankings* produzidos utilizando volume e a versão de volume normalizada entre subáreas, denominada ca -volume. É possível observar que o *ranking* produzido utilizando o volume de publicação é massivamente aperfeiçoado com o método de *ranking* entre subáreas proposto. De fato, ca -volume supera o tradicional índice volume de publicação com uma margem significativa a partir do corte k igual a 10 em diante.

De forma similar, a Figura 5.4 apresenta os valores de nDCG para os *rankings* gerados utilizando a quantidade de citações e a quantidade de citações depois de se

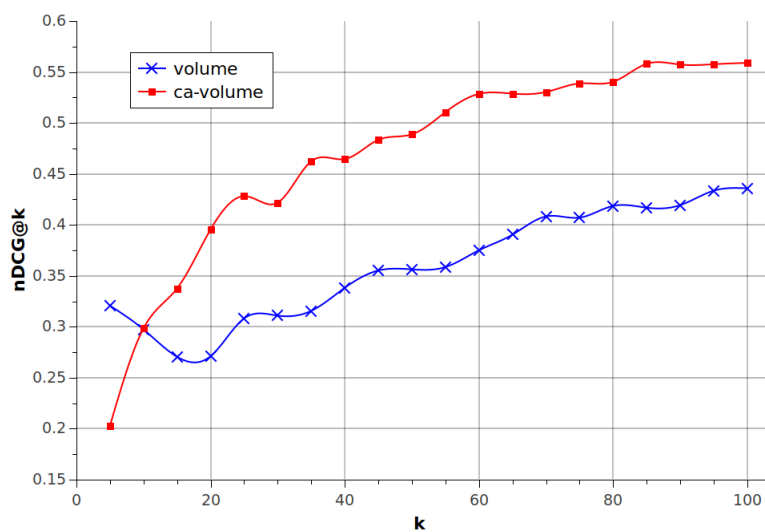


Figura 5.3. Comparação entre o *ranking* de pesquisadores utilizando o volume de publicação e o *ca-volume*.

aplicar o método proposto, denominado *ca-citação*. Nesse caso, é identificado um comportamento similar ao apresentado na Figura 5.3, embora apresente uma melhoria levemente inferior comparado com o *ca-volume*. O *ca-citação* apresenta uma melhora nos valores de nDCG para valores de corte k superiores a 25.

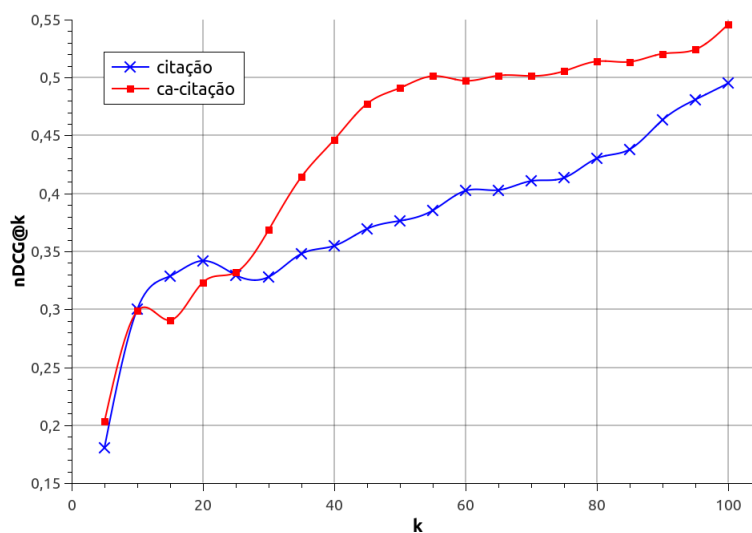
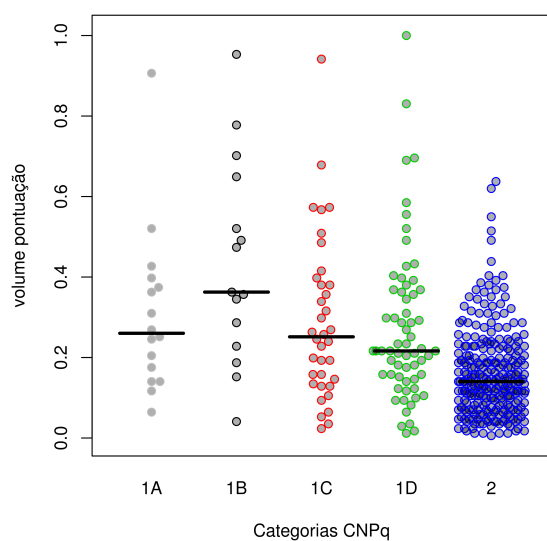
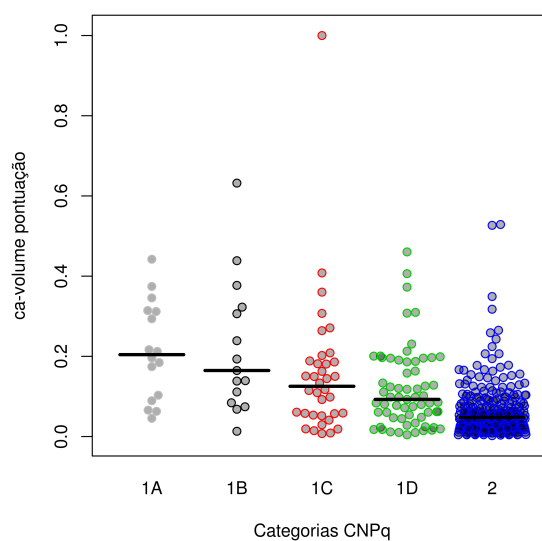
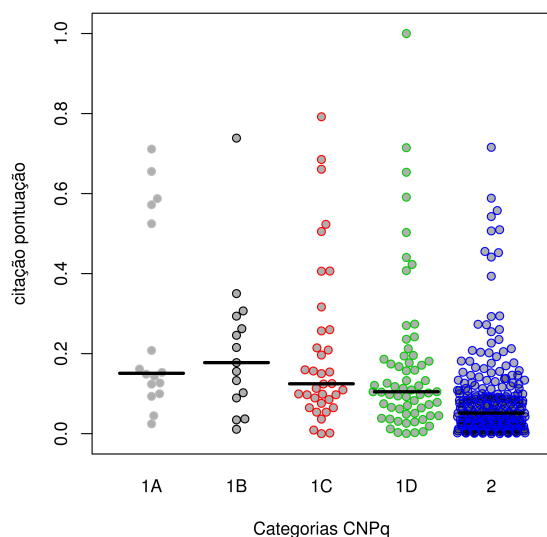


Figura 5.4. Comparação entre o *ranking* de pesquisadores utilizando o número de citação e o *ca-citação*.

Por fim, ainda com a intenção de demonstrar o desempenho da abordagem proposta para estimar volume de publicação e quantidade de citação entre pesquisadores de subáreas distintas, a Figura 5.5 mostra a dispersão dos valores produzidos usando tais



(a) distribuição dos valores de volume de produção

(b) distribuição dos valores de *ca*-volume

(c) distribuição dos valores de citação

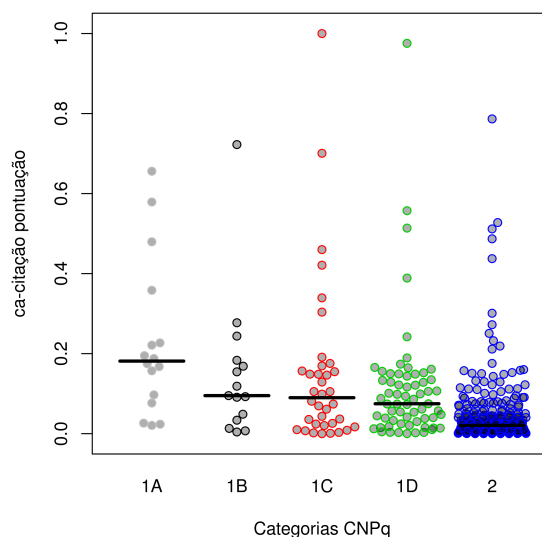
(d) distribuição dos valores de *ca*-citação

Figura 5.5. Distribuição de valores de (a) volume, (b) *ca*-volume, (c) citação e (d) *ca*-citação de acordo com os cinco níveis de relevância.

índices, antes e depois de aplicar o método de normalização entre subáreas. É possível observar que o método proposto ameniza a dispersão causada pela simples contagem do número de publicação e citação sem considerar possíveis especificidades de cada subárea. Nesse sentido, tanto *ca*-volume (Figura 5.5(b)) quanto *ca*-citação (Figura 5.5(d)) apresentam valor da mediana estritamente descendente a partir da categoria 1A à categoria 2, enquanto seus correspondentes globais (i.e., volume na Figura 5.5(a) e citação na Figura 5.5(c)) falham em alinhar corretamente as duas categorias de maior destaque, categorias 1A e 1B.

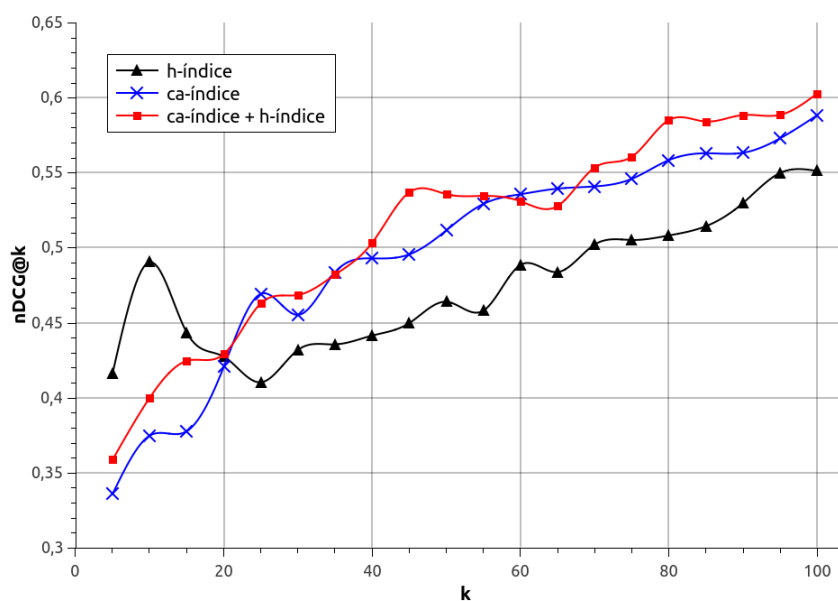


Figura 5.6. Comparação entre o *ranking* de pesquisadores com base no índice h , ca -índice e a combinação linear desses.

De forma geral, os resultados apresentados respondem a questão de pesquisa $Q1$, demonstrando o bom desempenho do método de *ranking* proposto em uma avaliação com pesquisadores de múltiplas subáreas de uma área específica. No que se refere aos índices de produtividade amplamente utilizados (i.e., volume de publicação, quantidade de citações e índice h), o método proposto consegue gerar um *ranking* que mais se aproxima ao *ranking* de referência produzido pelo CNPq.

5.4.2 Ranking utilizando múltiplos índices

Os resultados na seção anterior certificam a eficácia do método proposto para produzir *ranking* de pesquisadores de subáreas distintas em relação aos índices de produtividade originais. Por outro lado, o desempenho relativamente melhor obtido pelos índices globais nas primeiras posições do *ranking* sugere que tais índices fornecem evidências complementares que podem ser melhor exploradas na abordagem proposta. Para investigar tal hipótese, a questão de pesquisa $Q2$ é discutida analisando a combinação do índice h tradicional com o ca -índice. Tal combinação é linear com pesos iguais para ambos na combinação final. Como trabalho futuro é proposto a identificação automática dos pesos ótimos a partir dos dados disponíveis.

A Figura 5.6 mostra as curvas correspondentes ao índice h , ca -índice e à combinação desses. É possível observar que a combinação índice $h+ca$ -índice obteve um desempenho, pelo menos, tão eficiente quanto ao ca -índice (índice que teve o melhor de-

sempenho entre os dois utilizados) em termos de nDCG para quase todos os valores de corte k considerados. Esse resultado responde a questão de pesquisa $Q2$, demonstrando que existe escopo para melhorar o ca -índice, particularmente nas primeiras posições do *ranking*, combinando-o com outros índices globais.

5.4.3 Análise de falha

Essa seção analisa o *ranking* produzido utilizando o ca -índice, a versão do índice h comparável entre pesquisadores de subáreas distintas, de acordo com a classificação dos pesquisadores brasileiros de Ciência da Computação. Assim, a questão de pesquisa $Q3$ é respondida, apontando quais outros fatores podem impactar nos resultados gerados pela abordagem. Além disso, é discutido como o *ranking* produzido pelo ca -índice pode ser utilizado para identificar pesquisadores a serem selecionados para uma possível promoção ou rebaixamento de categoria (naturalmente, uma análise mais aprofundada é necessária).

Para facilitar tal comparação, a Tabela 5.1 apresenta uma matriz de confusão, organizando o universo de pesquisadores considerados nos experimentos de acordo com a classificação do CNPq e pelo ca -índice⁵. Em particular, cada célula c_{ij} na Tabela 5.1 apresenta a porcentagem de pesquisadores classificados pelo CNPq na categoria i e classificados pelo ca -índice na categoria j . Note que os valores percentuais são computados considerando *todos* pesquisadores, ao contrário de uma normalização apenas com os pesquisadores de um categoria específica. Em outras palavras, a linha 1A apresenta como o ca -índice ordenou os pesquisadores que estão atualmente na categoria 1A, segundo o CNPq. Por exemplo, o ca -índice classificou 0,75% dos pesquisadores definidos como 1A pelo CNPq como pertencentes à categoria 1B (conforme mostrado na primeira linha, segunda coluna).

A porcentagem de pesquisadores classificados corretamente (de acordo com CNPq) é apresentado na diagonal principal (total de 62%). Se todos os pesquisadores fossem corretamente classificados, a diagonal principal somaria 100%, e as demais células seriam iguais a 0. De forma geral, os valores da diagonal principal da matriz são relativamente altos, principalmente para os pesquisadores pertencentes à categoria 2. As discrepâncias apresentadas podem ser explicadas principalmente pelo fato de o CNPq utilizar outras fontes de evidências para classificar os pesquisadores, como número de estudantes orientados, contribuição para inovação, liderança de grupo de pesquisa e participação em comitês internacionais. Outro fator que pode influenciar

⁵A classificação do ca -índice é definida dividindo a distribuição de valores produzidos pelo ca -índice em cinco categorias com o mesmo tamanho das categorias correspondentes do CNPq, como retratada na Figura 3.1.

		ca-índice				
		1A	1B	1C	1D	2
CNPq	1A	1,01	0,75	0,75	0,5	1,01
	1B	0,75	0,75	0,5	1,51	0,25
	1C	0,75	0,75	1,76	3,02	2,76
	1D	1,01	1,01	1,76	3,52	8,54
	2	0,5	0,5	4,27	7,04	55,03

Tabela 5.1. Matriz de confusão para os *rankings* produzidos pelo *ca*-índice (colunas) e pela classificação do CNPq (linhas). As células destacadas mostram as discrepâncias entre a classificação dos pesquisadores das categorias 1 e 2. Os valores percentuais foram calculados com relação a *todos* os pesquisadores considerados.

nesse resultado é ano em que é feita a avaliação do CNPq para promover ou rebaixar pesquisadores de categoria. Como os dados utilizados foram de 2001 a 2011, e o CNPq realizou uma avaliação para promover pesquisadores em 2013, então possivelmente existem alguns pesquisadores com potencial para mudar de categoria que foram classificados de forma equivocada pelo método proposto. Essa análise é discutida na próxima seção.

Uma análise similar pode ser feita agrupando todos pesquisadores das subcategorias 1A-1D em uma ampla categoria 1. A principal diferença entre as categorias do CNPq 1 e 2 é que pesquisadores da categoria 1 ganham uma taxa de bancada além da bolsa. Portanto, a correta classificação dos pesquisadores nessas duas categorias pode ser útil para o CNPq por dois motivos: decidir quais pesquisadores devem receber o financiamento regular e quais devem receber também a taxa de bancada. Note que a distinção entre os pesquisadores das categorias 1 e 2 é uma tarefa que mais demanda tempo, pois a maioria dos pesquisadores são alocados na categoria 2. Nesse cenário de classificação binária, a abordagem proposta produz uma classificação correta para 75% dos pesquisadores (20% na categoria 1 e 55% na categoria 2), conforme a Tabela 5.2.

		ca-índice	
		Categoria 1	Categoria 2
CNPq	Categoria 1	20,10	12,56
	Categoria 2	12,31	55,03

Tabela 5.2. Matriz de confusão considerando os pesquisadores das categorias 1A-1D em uma ampla categoria 1.

Outra forma de interpretar os resultados da Tabela 5.1 é considerar que essa tabela resume a distribuição de bolsas de acordo com a produção dos pesquisadores.

Portanto, esses resultados podem ajudar no processo de tomada de decisão apontando os pesquisadores que potencialmente podem receber promoção ou serem rebaixados de categorias. Especificamente, todos pesquisadores acima da diagonal principal poderiam ser rebaixados para categoria 2; por outro lado, aqueles abaixo da diagonal principal poderiam ser considerados para receberem promoção para categoria 1. Por exemplo, a última linha da tabela representa os pesquisadores que foram classificados pelo *ca*-índice na categoria 1, entretanto estão classificados pelo CNPq na categoria 2. O perfil desses pesquisadores poderia ser melhor analisado para promoção à categoria 1. De forma semelhante, a coluna destacada na parte direita da tabela representa os pesquisadores classificados na categoria 2, mas estão atualmente na categoria 1 de acordo com CNPq. Tais pesquisadores são candidatos a serem rebaixados de categoria.

Normalmente, tais decisões não podem ser tomadas de forma automática, pois é necessário entender o porquê de tais discrepâncias apresentadas na Tabela 5.1. Por exemplo, os pesquisadores na linha 1A coluna 2 (pesquisadores 1A, segundo CNPq, mas classificados pelo *ca*-índice como categoria 2) podem ser aqueles pesquisadores que possuem o foco em outras atividades acadêmicas que não estão relacionadas a publicações. Por outro lado, os pesquisadores na linha 2 coluna 1A (classificados na categoria 2, segundo o CNPq, mas com *ca*-índice foram classificados na categoria 1A) podem ser aqueles que têm muitas publicações, mas possuem um perfil simples nas demais atividades acadêmicas (por exemplo, podem ainda não ter orientado aluno de doutorado).

5.4.4 Análise de promoção dos pesquisadores

Para responder a questão de pesquisa *Q4*, é analisado o comportamento do método diante das promoções concedidas pelo CNPq em 2013. Como mencionado na Seção 5.3, para avaliar o método proposto foi utilizado o ranking produzido pelo CNPq em 2012. No final de 2013, o CNPq realizou uma nova avaliação, promovendo cerca de 50 pesquisadores. Vale ressaltar que nessa avaliação nenhum pesquisador foi rebaixado de categoria e não houve perda de bolsa.

A Figura 5.7 mostra a distribuição da pontuação atribuída aos pesquisadores de acordo com o *ca*-índice. Nessa figura os pontos em vermelho mostram os pesquisadores que subiram um nível de acordo com a avaliação feita pelo CNPq. Ou seja, um ponto vermelho na categoria 2 mostra que esse pesquisador em 2012 estava na categoria 2, mas de acordo com a nova avaliação foi promovido para categoria 1D. O método conseguiu apontar alguns pesquisadores promovidos com produção destacada em relação aos demais de sua categoria, principalmente das categorias 1B, 1C e 2. Em

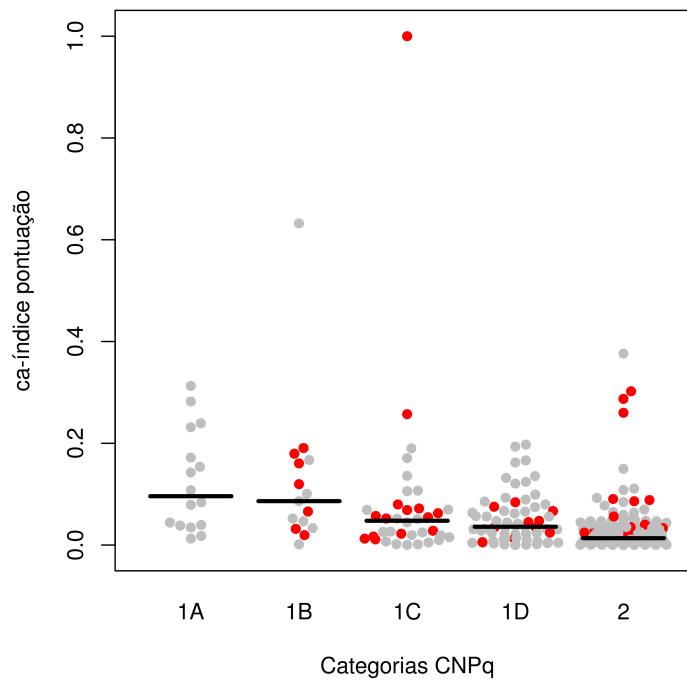


Figura 5.7. Pesquisadores promovidos de categoria segundo avaliação do CNPq realizada em 2013.

geral, a maior parte dos pesquisadores promovidos estão acima da mediana de pontuação de sua categoria, exceção ocorre com os pesquisadores promovidos à categoria *1C*. Como mencionado na análise de falha do método apresentada na seção anterior, para avaliar os pesquisadores são considerados diversos critérios, não somente a produção científica, assim tais critérios possivelmente foram considerados para promoção de tais pesquisadores.

Em suma, os pesquisadores com maiores valores de *ca*-índice em cada categoria são promovidos, apesar de alguns pesquisadores com baixa pontuação, principalmente da categoria *1D*, terem sido promovidos conforme mostrado na Figura 5.7.

Capítulo 6

Conclusões e trabalhos futuros

Nesta dissertação foi realizada a caracterização dos pesquisadores brasileiros com destacada produção acadêmica da área de Ciência da Computação financiados pelo CNPq. O perfil dos pesquisadores foi analisado considerando três dimensões: tempo de carreira, quantidade de alunos orientados, e volume de publicação e quantidade de citações. Além disso, é feita uma análise aprofundada das publicações dos pesquisadores de acordo com o impacto do veículo de publicação e a subárea de interesse do pesquisador. Diante de tais análises foi possível obter o retrato acadêmico dos pesquisadores de Ciência da Computação financiados pelo CNPq.

Os resultados desse estudo mostraram que o padrão de publicação do pesquisador está diretamente relacionado com a sua subárea de interesse. Diante disso, foi proposto um método de *ranking* de pesquisadores que considere as múltiplas subáreas de atuação do pesquisador. A abordagem proposta foi comparada com índices de produtividade amplamente utilizados como o índice h , produzindo melhores resultados de acordo com *ranking* do CNPq.

Em outras palavras, as principais contribuições desta dissertação podem ser resumidas da seguinte maneira:

- **Análise dos pesquisadores de Ciência da Computação.** No Capítulo 3 é apresentada uma análise dos pesquisadores de Ciência da Computação financiados pelo CNPq. Esse estudo caracterizou os pesquisadores segundo três aspectos: tempo de carreira, quantidade de alunos orientados, volume de publicação e quantidade de citações. Além disso, a produção científica dos pesquisadores é avaliada de acordo com impacto do veículo de publicação e a subárea de interesse do pesquisador. Com esse estudo foi possível mostrar o perfil em relação às atividades acadêmicas de tais pesquisadores e como o desempenho dos pesquisadores nas

categorias do CNPq se diferenciam.

- **Premissas de avaliação de pesquisadores de subáreas distintas.** Para propiciar uma avaliação imparcial de pesquisadores com perfis de subáreas distintas, na Seção 4.1 são propostas três premissas consideradas em tais avaliações: *Pluralidade* (produtividade do pesquisador deve ser medida em todas as subáreas de atuação do pesquisador), *Diversidade* (perfil de cada subárea de pesquisa deve ser considerado na avaliação de produtividade do pesquisador) e *Igualdade* (as subáreas de pesquisa devem ser consideradas igualmente importantes e merecedoras de mérito científico).
- **Método de *ranking* de pesquisadores de múltiplas subáreas.** Inspirando em abordagens de agregação de *ranking*, no Capítulo 4 é proposto um método de *ranking* de pesquisadores com diversas subáreas de atuação. A abordagem proposta é genérica e produz versões comparáveis entre subáreas de índices globais, como índice h , quantidade de citações e volume de publicação. A avaliação realizada no Capítulo 5 mostrou que o método proposto produz melhores resultados de acordo com *ranking* oficial produzido pelo CNPq. Nessa avaliação também foram identificados possíveis fatores que influenciam no desempenho do método, sendo que tais fatores podem ser utilizados para melhorar a abordagem proposta.
- Os resultados desta dissertação também foram publicados em [Lima et al., 2013], e foram reorganizados em um artigo submetido para o periódico *Scientometrics*¹.

Finalmente, como possibilidades de trabalhos futuros citam-se:

- **Melhoria do método proposto de *ranking* de pesquisadores com múltiplas subáreas de atuação.** Como mostrado na avaliação dos pesquisadores na Seção 5.4.3, diversos outros fatores são utilizados na avaliação realizada pelo CNPq (e.g., orientação de alunos, liderança de grupos de pesquisa, inserção internacional, etc). Assim, tais fatores poderiam também ser incorporados no método proposto a fim de produzir um *ranking* mais próximo do ideal.
- **Identificação de forma automática de pesos para a combinação de índices globais com a abordagem proposta.** Na Seção 5.4.2 o método proposto é combinado linearmente com índice h . Em trabalhos futuros poderia ser aplicado algum método para identificar pesos ótimos a partir do conjunto de dados disponível, possivelmente gerando melhores resultados.

¹Scientometrics: <http://link.springer.com/journal/11192>

- **Aplicar o método proposto de *ranking* em outras áreas científicas.** Neste trabalho, o método proposto foi avaliado na área de Ciência da Computação. Entretanto, nada impede que o método seja aplicado a outras áreas científicas. Assim, uma possível avaliação do método poderia cobrir outras áreas.
- **Análise aprofundada por subárea.** Para gerar o *ranking* global apresentado na Seção 4.2, é necessário gerar os *rankings* de cada subárea. Em trabalhos futuros, tais *rankings* poderiam ser melhor explorados para analisar o desempenho dos pesquisadores CNPq em cada subárea. Assim, seria possível perceber se os pesquisadores das categorias superiores do CNPq possuem também melhor desempenho quando comparados separadamente por subárea.

Referências Bibliográficas

- Abramo, G.; D'Angelo, C. A. & Di Costa, F. (2011). National research assessment exercises: the effects of changing the rules of the game during the game. *Scientometrics*, 88(1):229--238.
- Almeida, E. & Guimarães, J. (2013). Brazil's growing production of scientific articles – how are we doing with review articles and other qualitative indicators? *Scientometrics*, pp. 1–29.
- Althouse, B. M.; West, J. D.; Bergstrom, C. T. & Bergstrom, T. (2009). Differences in impact factor across fields and over time. *JASIST*, 60(1):27--34.
- Arruda, D.; de Lima Bezerra, F.; Neris, V. A.; Toro, P. R. D. & Wainer, J. (2009). Brazilian computer science research: Gender and regional distributions. *Scientometrics*, 79(3):651–665.
- Barbosa, S. D. J. & de Souza, C. S. (2011). Are HCI researchers an endangered species in Brazil? *ACM Interactions Magazine*, 18(3):69--71.
- Bornmann, L. & Daniel, H.-D. (2009). Universality of Citation Distributions - A Validation of Radicchi et al.'s Relative Indicator $cf = c/c_0$ at the Micro Level Using Data From Chemistry. *JASIST*, 60(8):1664--1670.
- Bornmann, L. & Marx, W. (2014). How to evaluate individual researchers working in the natural and life sciences meaningfully? a proposal of methods based on percentiles of citations. *Scientometrics*, 98(1):487–509.
- Bornmann, L.; Mutz, R. & Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *JASIST*, 59(5):830–837.
- Callan, J. (2000). Distributed information retrieval. Em Croft, W. B., editor, *Advances in Information Retrieval*, capítulo 5, pp. 127--150. Kluwer Academic Publishers.

- Clarke, C. L. A.; Craswell, N. & Soboroff, I. (2009). Overview of the TREC 2009 Web track. Em *Procs. of TREC*, Gaithersburg, MD, USA.
- Claro, J. & Costa, C. A. V. (2011). A made-to-measure indicator for cross-disciplinary bibliometric ranking of researchers performance. *Scientometrics*, 86(1):113–123.
- Collazo-Reyes, F. (2014). Growth of the number of indexed journals of latin america and the caribbean: the effect on the impact of each country. *Scientometrics*, 98(1):197–209.
- Cota, R. G.; Ferreira, A. A.; Nascimento, C.; Gonçalves, M. A. & Laender, A. H. F. (2010). An unsupervised heuristic-based hierarchical method for name disambiguation in bibliographic citations. *JASIST*, 61(9):1853–1870.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1):131–152.
- Ferreira, A. A.; Gonçalves, M. A. & Laender, A. H. F. (2012). A brief Survey of Automatic Methods for Author Name Disambiguation. *SIGMOD Record*, 41(2):15–26.
- Freire, V. P. & Figueiredo, D. R. (2011). Ranking in collaboration networks using a group based metric. *J. Braz. Comp. Soc.*, 17(4):255–266.
- Glänzel, W.; Leta, J. & Thijs, B. (2006). Science in Brazil. Part 1: A macro-level comparative study. *Scientometrics*, 67(1):67–86.
- Glänzel, W. & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3):357–367.
- Helene, A. & Ribeiro, P. (2011). Brazilian scientific production, financial support, established investigators and doctoral graduates. *Scientometrics*, 89(2):677–686.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS*, 102(46):16569–16572.
- Hyndman, R. J. & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, 50:361–365.
- Ibáñez, A.; Larrañga, P. & Bielza, C. (2013). Cluster methods for assessing research performance: exploring spanish computer science. *Scientometrics*, 97(3):571–600.

- Järvelin, K. & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.*, 20(4):422--446.
- Kato, M. & Ando, A. (2013). The relationship between research performance and international collaboration in chemistry. *Scientometrics*, 97(3):535–553.
- Kutlar, A.; Kabasakal, A. & Ekici, M. (2013). Contributions of turkish academicians supervising phd dissertations and their universities to economics: an evaluation of the 1990-2011 period. *Scientometrics*, 97(3):639–658.
- Laender, A. H. F.; de Lucena, C. J. P.; Maldonado, J. C.; de Souza e Silva, E. & Ziviani, N. (2008). Assessing the research and education quality of the top Brazilian Computer Science graduate programs. *SIGCSE Bulletin*, 40(2):135–145.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38(21):201–221.
- Lane, J. (2010). Let's make science metrics more scientific. *Nature*, 464(7288):488--489.
- Lee, D.; Kang, J.; Mitra, P.; Giles, C. L. & On, B.-W. (2007). Are your citations clean? *Commun. ACM*, 50(12):33–38.
- Leskovec, J.; Backstrom, L.; Kumar, R. & Tomkins, A. (2008). Microscopic evolution of social networks. Em *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pp. 462--470, New York, NY, USA. ACM.
- Leta, J.; Glänzel, W. & Thijs, B. (2006). Science in Brazil. Part 2: Sectoral and institutional research profiles. *Scientometrics*, 67(1):87–105.
- Leydesdorff, L. (2012). Alternatives to the journal impact factor: I3 and the top-10% (or top-25%?) of the most-highly cited papers. *Scientometrics*, 92(2):355–365.
- Lima, H.; Silva, T. H.; Moro, M. M.; Santos, R. L.; Meira, Jr., W. & Laender, A. H. (2013). Aggregating productivity indices for ranking researchers across multiple areas. Em *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '13, pp. 97--106.
- Lundberg, J. (2007). Lifting the crown - citation z-score. *J. Informetrics*, 1(2):145–154.
- Michels, C. & Schmoch, U. (2014). Impact of bibliometric studies on the publication behaviour of authors. *Scientometrics*, 98(1):369–385.

- Miller, J.; Coble, K. H. & Lusk, J. L. (2013). Evaluating top faculty researchers and the incentives that motivate them. *Scientometrics*, 97(3):519–533.
- Moksony, F.; Hegedüs, R. & Császár (2013). Rankings, research styles, and publication cultures: a study of american sociology departments. *Scientometrics*, online.
- Oliveira, E. A.; Colosimo, E. A.; Martelli, D. R.; Quirino, I. G.; Oliveira, M. C.; Lima, L. S.; Simões E Silva, A. C. & Martelli-Júnior, H. (2012). Comparison of Brazilian researchers in clinical medicine: are criteria for ranking well-adjusted? *Scientometrics*, 90(2):429–443.
- Pallis, G.; Zeinalipour-Yazti, D. & Dikaiakos, M. D. (2011). Online social networks: Status and trends. Em *New Directions in Web Data Management 1*, pp. 213–234.
- Podlubny, I. (2005). Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*, 64(1):95–99.
- Radicchi, F. & Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, 97(3):627–637.
- Radicchi, F.; Fortunato, S. & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *PNAS*, 105:17268–17272.
- Ribas, S.; Ribeiro-Neto, B. A.; de Souza e Silva, E. & Ziviani, N. (2013). R-score: Reputation-based scoring of research groups. *CoRR*, abs/1308.5286.
- Roos, D. H.; Calabro, L.; Jesus, S. L.; Souza, D. O.; Barbosa, N. V. & Rocha, J. a. B. T. (2014). Brazilian scientific production in areas of biological sciences: a comparative study on the modalities of full doctorate in Brazil or abroad. *Scientometrics*, 98(1):415–427.
- Sangwal, K. (2013). Some citation-related characteristics of scientific journals published in individual countries. *Scientometrics*, 97(3):719–741.
- Shelton, R. & Lewison, G. (2013). Scientific collaboration as a window and a door into north korea. *Scientometrics*, 97(1):3–11.
- Sugimoto, C. R. & Cronin, B. (2012). Biobibliometric profiling: An examination of multifaceted approaches to scholarship. *JASIST*, 63(3):450–468.
- Sune Lehmann, Andrew D. Jackson, B. E. L. (2006). Measures for measures. *Nature*, 444:1003–1004.

- Torrise, B. (2013). A multidimensional approach to academic productivity. *Scientometrics*, online.
- Vanecek, J. (2014). The effect of performance-based research funding on output of r&d results in the czech republic. *Scientometrics*, 98(1):657–681.
- Veloso, A.; Meira, Jr., W.; Gonçalves, M. & Zaki, M. (2007). Multi-label lazy associative classification. Em *Procs. of ECML/PKDD*, pp. 605–612.
- Wagner, C. S.; Roessner, J. D.; Bobb, K.; Klein, J. T.; Boyack, K. W.; Keyton, J.; Rafols, I. & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (idr): A review of the literature. *Journal of Informetrics*, 5(1):14 – 26.
- Wainer, J.; Eckmann, M.; Goldenstein, S. & Rocha, A. (2013). How productivity and impact differ across computer science subareas. *Commun. ACM*, 56(8):67–73.
- Wainer, J.; Goldenstein, S. & Billa, C. Z. (2011). Invisible work in standard bibliometric evaluation of computer science. *Commun. ACM*, 54(5):141–146.
- Wainer, J.; Xavier, E. C. & de Lima Bezerra, F. (2009). Scientific production in computer science: A comparative study of brazil and other countries. *Scientometrics*, 81(2):535–547.
- Waltman, L.; Eck, N. J.; Leeuwen, T. N.; Visser, M. S. & Raan, A. F. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics*, 87(3):467–481.
- Waltman, L.; van Eck, N. J.; van Leeuwen, T. N.; Visser, M. S. & van Raan, A. F. J. (2010). Towards a new crown indicator: Some theoretical considerations. *CoRR*, abs/1003.2167.