

# Revista Eletrônica de Sistemas de Informação

## ISSN 1677-3071

V. 14, n. 1

jan-abr 2015 - Edição Temática sobre Análise de Redes Sociais e Mineração

doi:10.21529/RESI.2015.1401

### Sumário

#### Editorial

##### EDITORIAL

*Jonice Oliveira*

#### BrASNAM

##### ANÁLISE DA EVOLUÇÃO DAS RELAÇÕES DE COAUTORIA NOS PROGRAMAS DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO NO BRASIL

*Luciano A. Digiampietri, Jesús P. Mena-Chalco, Gabriela S. Silva, Leonardo B. Oliveira, Jamison J. S. Lima, Ana Paula Malheiro, Dania Meira*

##### ANÁLISE COMPARATIVA DA PRODUTIVIDADE DOS PARES ORIENTADOR-ORIENTADO EM CIÊNCIA DA COMPUTAÇÃO

*Karina Valdivia-Delgado, Esteban Fernandez-Tuesta, Luciano Digiampietri, Rogério Mugnaini, Jesús P. Mena-Chalco, José J. Pérez-Alcázar*

##### MINERANDO PUBLICAÇÕES CIENTÍFICAS PARA ANÁLISE DA COLABORAÇÃO EM COMUNIDADES DE PESQUISA – O CASO DA COMUNIDADE DE SISTEMAS DE INFORMAÇÃO

*Renata Mendes de Araujo, Brunno Silveira, Thiago Muramatsu, Kate Revoredo*

##### APRENDIZADO DE MÁQUINA PARA ROTULAÇÃO AUTOMÁTICA DE USUÁRIOS DE UMA REDE SOCIAL ACADÊMICA

*Bruno Vicente Alves de Lima, Vinicius Ponte Machado, Lucas Araújo Lopes*



Este trabalho está licenciado sob uma [Licença Creative Commons Attribution 3.0](http://creativecommons.org/licenses/by/3.0/).

ISSN: 1677-3071

Esta revista é (e sempre foi) eletrônica para ajudar a proteger o meio ambiente, mas, caso deseje imprimir esse artigo, saiba que ele foi editorado com uma fonte mais ecológica, a *Eco Sans*, que gasta menos tinta.

*This journal is (and has always been) electronic in order to be more environmentally friendly. Now, it is desktop edited in a single column to be easier to read on the screen. However, if you wish to print this paper, be aware that it uses Eco Sans, a printing font that reduces the amount of required ink.*

# ANÁLISE COMPARATIVA DA PRODUTIVIDADE DOS PARES ORIENTADOR-ORIENTADO EM CIÊNCIA DA COMPUTAÇÃO

## COMPARATIVE ANALYSIS OF THE PRODUCTIVITY OF ADVISOR-ADVISEE PAIRS IN COMPUTER SCIENCE

(artigo submetido em março de 2013)

**Esteban F. Tuesta**

Escola de Artes, Ciências e Humanidades -  
Universidade de São Paulo (EACH-USP)  
tuesta@usp.br

**Luciano A. Digiampietri**

Escola de Artes, Ciências e Humanidades -  
Universidade de São Paulo (EACH-USP)  
jamison.lima@usp.br

**Jesús P. Mena-Chalco**

Centro de Matemática, Computação e  
Cognição – Univ. Federal do ABC (UFABC)  
jesus.mena@ufabc.edu.br

**Karina V. Delgado**

Escola de Artes, Ciências e Humanidades -  
Universidade de São Paulo (EACH-USP)  
kvd@usp.br

**Rogério Mugnaini**

Escola de Artes, Ciências e Humanidades -  
Universidade de São Paulo (EACH-USP)  
mugnaini@usp.br

**José J. Pérez-Alcázar**

Escola de Artes, Ciências e Humanidades -  
Universidade de São Paulo (EACH-USP)  
jperez@usp.br

### ABSTRACT

*The increasing involvement of graduate students in scientific communication has been the subject of studies in many countries that have been discussing the role of doctoral studies in the research career. In this article we are interested on the study of the academic relationship between the advisor and the advisee for a group of PhD pairs in the Computer Science area, with curricula registered in the Lattes Platform. We analyze the main characteristics of the group and their coauthor relationships. In terms of co-authorship: (i) we observe that the duration that is established between advisor and advisee can extrapolate the formal period of supervision, and (ii) we show that the duration of the collaboration is correlated with both the number of articles published by the advisee and the ones published by the advisor. In this work we also present the advisee profile for researchers that worked in some of the graduate courses in Computer Science in Brazil in the triennium 2007-2009.*

*Keywords: Lattes platform; social network; co-authorship network; advisor-advisee relationship.*

### RESUMO

A crescente participação dos pós-graduandos na comunicação científica das diversas áreas da ciência tem sido tema de estudos em diversos países que vêm discutindo o papel do doutoramento na carreira de pesquisa. Neste artigo é estudada a relação acadêmica temporal existente entre o aluno orientado e seu orientador para o conjunto de pares de doutores da área de Ciência da Computação com currículos cadastrados na plataforma Lattes. Analisamos as principais características do grupo e suas relações de coautoria. Em termos de coautoria: (i) observamos que o tempo de duração da parceria que se estabelece entre orientador e orientado pode extrapolar o período formal de orientação e (ii) mostramos que a duração do tempo de colaboração correlaciona-se com o número de artigos em periódicos publicados pelo orientado e o número de artigos publicados pelo orientador. Neste trabalho também apresentamos o perfil dos orientados para pesquisadores que atuaram como docentes de algum dos cursos de pós-graduação em Ciência da Computação no Brasil no triênio 2007-2009.

Palavras-chave: plataforma Lattes; rede social; rede de coautoria; relação orientador-orientado.

## 1 INTRODUÇÃO

A política científica brasileira conta com um amplo aparato institucional que abrange os Programas de Pós-Graduação de todo o país, consubstanciando-se em um vultoso processo de avaliação trienal, que reuniu em 2010 cerca de 900 consultores, com vistas à avaliação de 2.900 programas de pós-graduação ao longo de quatro semanas (AMARAL, 2010). Dentre os aspectos analisados, a produtividade científica é aferida pelo número de publicações em periódicos de qualidade científica reconhecida. Portanto, espera-se que o aluno pós-graduando seja inserido, desde seu ingresso no Programa, em um ciclo de pesquisa cujo resultado culmine na publicação de artigos. Dados sobre a formação de doutores no Brasil mostram que desde o ano 1981 até 2009 se formaram 96.782 doutores<sup>1</sup>.

Um estudo recente sobre a produtividade dos alunos do Programa de Pós-Graduação do Instituto de Pesquisas Energéticas e Nucleares, em seus mais de trinta anos de existência, mostra que o tempo para publicação de artigos relacionados à pesquisa de tese, que consiste de uma média da diferença entre os anos de publicação dos artigos e o ano de defesa da tese, vem diminuindo ao longo dos últimos dez anos (MUGNAINI *et al.*, 2011).

Essa crescente participação dos pós-graduandos na comunicação científica das diversas áreas da ciência tem sido tema de estudos em diversos países que vêm discutindo o papel do doutoramento na carreira de pesquisa, dentre eles: França (SALMI *et al.*, 2001), Croácia (FRKOVIC *et al.*, 2003), Brasil (RAMOS *et al.*, 2009, SACARDO e HAYASHI, 2011); e Peru (ARRIOLA-QUIROZ *et al.*, 2010) - a maioria deles em áreas de Ciência da Saúde -, e nos Estados Unidos (LEE, 2000, ANWAR, 2004, MALLETTE, 2006) e Canadá (LARIVIERE, 2010) - em áreas diversas. Geralmente estes estudos consistem do levantamento dos nomes dos autores de teses e a busca da respectiva produção científica em bases de dados de periódicos. Alguns deles utilizam o nome do orientador como critério para decisão sobre o artigo ter derivado da pesquisa doutoral. Outro aspecto é o tempo de publicação que se considera para levantamento dos artigos publicados: por exemplo: cinco anos antes e após a defesa da tese (MUGNAINI *et al.*, 2011).

Existem ainda estudos qualitativos que analisam as relações interpessoais entre orientador-orientado e as consideram importantes para o sucesso do projeto de doutorado. O instrumento utilizado para mensurar esta relação é na maior parte deles, um questionário (MAINHARD *et al.*, 2009). Outros pesquisadores (GOLDE, 2000, KAM, 1997, MARSH *et al.*, 2002, MCALPINE e NORTON, 2006) também indicam a importância dessa relação.

---

<sup>1</sup> <http://lattes.cnpq.br/painelLattes/evolucaoformacao/>

Outros trabalhos que estudam a relação acadêmica orientador-orientado na área de computação na literatura recente usam ferramentas de mineração de dados e aprendizado computacional para inferir a existência dessa relação, uma vez que dispõem apenas de conjuntos/listas de produções bibliográficas publicadas, e as informações sobre as relações estão escondidas entre as outras informações não relevantes (WU *et al.*, 2007, WU *et al.*, 2010, WANG *et al.*, 2010, WANG *et al.*, 2012).

O objetivo deste trabalho consistiu do estudo da relação temporal entre orientador e orientado para os dados de um conjunto de pesquisadores doutores da área de Ciência da Computação, com a finalidade de verificar se a duração dessa parceria influencia na produtividade do orientando. Este conjunto de pesquisadores encontrava-se cadastrado na Plataforma Lattes (uma das plataformas mais completas de currículos de pesquisadores afiliados a instituições brasileiras ou que têm relação com instituições brasileiras). A vantagem da utilização dos dados extraídos da Plataforma Lattes reside em que a relação orientador-orientado está identificada nessa base.

A contribuição principal deste trabalho sobre as relações acadêmicas orientador-orientado é a identificação do período e a sua influência na produtividade do orientado. Isto é, se o tempo de duração da coautoria estabelecida entre orientador e orientado (i) extrapola o período formal de orientação, e (ii) correlaciona-se com o número de produções bibliográficas do aluno orientado. O presente artigo revisita e estende o trabalho de Tuesta *et al.* (2012) de forma a analisar, no período 2007-2009, a diferença na produção acadêmica entre o grupo de orientados que atuaram como docentes de algum dos cursos de pós-graduação em Ciência da Computação no Brasil; e o grupo de orientados que não formaram parte de nenhum curso de pós-graduação.

O artigo está organizado da seguinte forma. A Seção 2 descreve o processo metodológico que inclui a seleção da amostra e as definições das variáveis utilizadas. A Seção 3 apresenta os resultados e as principais análises realizadas. A Seção 4 discute os resultados obtidos para o trabalho. As conclusões e considerações finais são apresentadas na Seção 5.

## 2 MATERIAL E MÉTODOS

Esta seção descreve o procedimento para obtenção da amostra dos dados obtidos de currículos da Plataforma Lattes, as perguntas de pesquisa e as variáveis usadas.

### 2.1 PROCEDIMENTO PARA OBTENÇÃO E DESCRIÇÃO DA AMOSTRA

Os dados analisados neste artigo foram extraídos dos currículos da Plataforma Lattes<sup>2</sup>. Segundo informações disponibilizadas pelo CNPq<sup>3</sup> se

---

<sup>2</sup> <http://lattes.cnpq.br/>

formaram 1.746 doutores em Ciência da Computação entre os anos 1981 e 2009. Foram também identificados dentro desse grupo os 889 currículos pertencentes aos docentes do corpo permanente dos 45 programas de pós-graduação acadêmica (que possuem doutorado) em computação presentes na avaliação do triênio 2007-2009, realizada pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Isto foi feito por meio de um processo semiautomático, considerando os *Cadernos de Indicadores*.

Os currículos foram baixados utilizando-se o scriptLattes (MENA-CHALCO e CESAR-JR, 2009), tabulados e salvos em um banco de dados por um *parser* que discriminou cada um dos itens do currículo. Cada currículo foi dividido segundo três seções: formação do pesquisador, orientações e publicações em periódicos. Os dados selecionados foram as publicações por ano em periódicos, ano de início de doutorado, ano de conclusão de doutorado, número de coautores de cada publicação. Além disso, para cada par orientador-orientado, dispõe-se das publicações em conjunto assim como o número de coautores destas publicações. Os anos de todas as publicações não são superiores a 2010 nem inferiores a 1981 e os anos de conclusão de doutorado não são superiores a 2009.

As relações de orientação foram estabelecidas por meio de uma verificação dupla entre os currículos analisados, sendo consideradas relações de orientação apenas aquelas que apareciam nos currículos do orientado e do orientador, da seguinte forma: dentro da formação/titulação de doutorado no currículo do orientado, e dentro da lista de orientações de doutorado, no currículo do orientador.

Utilizando estes critérios, foram identificados 1.166 pares orientador-orientado que atuaram em *Ciência da Computação*. Dos 1.166 pares, 141 pares são pertencentes ao grupo dos docentes do corpo permanente de algum dos 45 programas de pós-graduação em Ciência da Computação no triênio de 2007-2009, que configurou o conjunto dos *orientados com vínculo na pós-graduação*, denominado *Grupo A*. Os outros 1.025 pares constituíram o conjunto dos *orientados sem vínculo na pós-graduação*, denominado *Grupo B*.

Finalmente, os currículos foram selecionados e as relações de coautoria identificadas, para então comparar os títulos das publicações e a lista de coautores, utilizando um algoritmo de casamento inexato de *strings*.

Foram considerados para análise, apenas os artigos em periódicos, dos quais 1.776 corresponderam a trabalhos em conjunto (orientador-orientado) de Computação, 12.562 artigos escritos pelos orientadores e 4.298 pelos orientados. É importante esclarecer que, na contagem dos artigos publicados individualmente pelos orientadores e orientados estão considerados os 1.776 artigos publicados em conjunto<sup>4</sup> (representando 14,14% dos artigos escritos pelos orientadores e 41,32% dos escritos pelos orientados).

---

<sup>3</sup> <http://lattes.cnpq.br/painelLattes/evolucaoformacao/>

<sup>4</sup> Para análise dos dados, foi utilizado o pacote estatístico R-project versão 2.14.2 (<http://www.r-project.org/>)

No Grupo A, 2.333 artigos foram escritos pelos orientadores e 959 pelos orientados, dos quais foram identificados 389 artigos em conjunto, representando 16,67% dos trabalhos escritos pelos orientadores e 40,56% dos orientados.

No Grupo B, 10.229 artigos foram escritos pelos orientadores e 3.339 pelos orientados dos quais foram identificados 1.387 artigos em conjunto, representando 13,56% dos trabalhos dos orientadores e 41,54% dos orientados. Pode-se observar no Grupo A que os orientados têm uma produção total equivalente a 41,1 % da produção total dos orientadores e no Grupo B equivalente a 32,64%.

## 2.2 MÉTODO

As seguintes perguntas nortearam as análises, assim como a verificação de diferença entre os grupos considerados:

- Qual a duração do doutorado?
- Qual a proporção de orientados que não publicaram com seus respectivos orientadores?
- A duração do tempo de relacionamento pode extrapolar o período formal de orientação?
- Existem diferenças entre as distribuições do tempo até a primeira publicação com e sem o orientador?
- Qual o número total de publicações conjuntas?
- O tempo de relacionamento tem alguma influência no número total de publicações do orientado?

No Quadro 1 foram descritas as variáveis usadas e o seu propósito. Note que as variáveis estão relacionadas com questões apresentadas anteriormente.

Feita a análise descritiva das variáveis, foi utilizado o coeficiente de correlação de Spearman para calcular a relação entre os anos de relacionamento e o número total de publicações do orientado (Seção 3.2), e também com o número total de publicações conjuntas. O coeficiente de correlação de Spearman ( $\rho$  de Spearman) é uma medida não paramétrica que permite mensurar a força de associação entre duas variáveis. Quando existe uma relação muito forte, o valor de  $\rho$  estará próximo de 1 ou -1, conforme a associação seja positiva ou negativa, e um valor de  $\rho$  igual a zero ou próximo de zero indica a falta de associação. Diferencia-se do coeficiente de correlação de Pearson por não exigir normalidade das variáveis, nem linearidade da associação entre elas. O teste de hipótese sobre o coeficiente usa procedimentos assintóticos (MYERS e WELL, 2003).

Quadro 1. Descrição das variáveis usadas

Variável	Descrição	Propósito
Duração formal do doutorado	Definida como a diferença entre o ano do fim do doutorado e o ano de início.	Verificar se existem ou não diferenças para cada conjunto considerado.
Anos de relacionamento ou duração da colaboração	Definida como a diferença entre o ano da última publicação em conjunto e o ano de início do doutorado e quando não existem publicações do par, esta variável foi definida como a diferença entre o ano de conclusão e o ano de início do doutorado. Note que este período é mínimo uma vez que após a última publicação o par pode ainda estar trabalhando.	Verificar se existem ou não diferenças entre os grupos considerados e se este período extrapola o período formal de orientação. Foi calculado também o coeficiente de correlação de Spearman entre anos de relacionamento e o número total de publicações.
Tempo transcorrido até a primeira publicação	Esta variável considera o tempo transcorrido desde o fim do doutorado até a primeira publicação do orientado com e sem a participação do orientador. Valores negativos indicam publicações antes do fim do doutorado e valores positivos após este período.	Comparar a distribuição do tempo transcorrido até a primeira publicação com e sem a participação do orientador para cada um dos conjuntos acima citados.
Número de publicações em conjunto com o orientador	É definida como a quantidade de artigos produzidos pelo par em um período de tempo.	Comparar as médias e medianas para os grupos considerados. Esta comparação também é feita no caso da ausência de publicações conjuntas, isto é, os orientados não publicaram com seus respectivos orientadores.
Número total de publicações	Para todos os elementos na amostra, foi calculada a soma dos artigos publicados em periódicos.	Comparar as médias e medianas para cada um dos três subconjuntos.
Número de coautores para as publicações conjuntas	Para cada publicação em que apareçam o orientador e o orientado foi contado o número de coautores.	Comparar a média do número de coautores para cada par que pertence aos grupos considerados.

Fonte: elaborado pelos autores

Além disso, foram apresentadas as redes de coautoria e de orientações para cada grupo estudado. As redes de coautoria acadêmica e as redes de citações são classes de redes sociais que têm sido estudadas principalmente com a finalidade de determinar a estrutura das colaborações entre pesquisadores (MENA-CHALCO e CESAR JR., 2011) e avaliar o *status* dos pesquisadores (POBLACIÓN *et al.*, 2009). A rede de coautoria, diferentemente da rede de citações, implica em um relacionamento temporal e social entre os autores (LIU *et al.*, 2005). Uma subclasse especial da rede de coautorias consiste na rede formada pelos orientadores e seus alunos orientados de doutorado (chamada de rede de orientações), em que existe um compromisso formal entre os indivíduos de produzirem algum produto de natureza científica.

### 3 RESULTADOS

Nesta seção são apresentados os resultados obtidos da análise descritiva de cada variável, a análise de correlação e as redes de coautoria e orientação.

#### 3.1 ANÁLISE DESCRITIVA DA REDE

Na amostra total observaram-se 16.450 artigos, dos quais 916 (5,6%) foram escritos por um único autor e 1.776 artigos foram escritos pelos pares orientador-orientado (10,8%). O número médio de publicações dos orientados e dos orientadores de Computação foi 3,68 e 10,77, respectivamente. Para o Grupo A, foi 6,8 e 16,55, respectivamente, e para o Grupo B foi de 3,26 e 9,98, respectivamente.

A Tabela 1 apresenta a mediana, a média e o desvio padrão de cada variável analisada, contrapondo-se os grupos considerados.

Tabela 1. Medidas descritivas

	Computação			Grupo A			Grupo B		
	Mediana	Média	Desvio Padrão	Mediana	Média	Desvio Padrão	Mediana	Média	Desvio Padrão
Duração do doutorado (em anos)	4	4,41	1,78	4	4,27	1,24	4	4,43	1,17
Tempo de relacionamento do par (em anos)	5	5,58	2,96	7	7,27	4,04	5	5,35	2,7
Tempo até a primeira publicação conjunta com o orientador desde o fim do doutorado	0	-0,3	2,45	0	-0,11	2,62	0	-0,33	2,42
Tempo até a primeira publicação sem o orientador desde o fim do doutorado	0	0,26	3,54	2	1,8	3,52	0	-0,03	3,47
Número total de publicações em conjunto com o orientador	1	1,52	2,23	2	2,76	2,76	1	1,35	2,09
Número médio de coautores para todas as publicações em conjunto de cada par	3,5	3,89	1,99	3,86	4,03	1,7	3,33	3,86	2,04

Fonte: elaborada pelos autores

### 3.1.1 Tempo para concluir o doutorado

Quanto ao tempo que um aluno demora a concluir seu doutorado, observou-se que a moda e a mediana foram de 4 anos para todos os grupos, ao passo que a média para o Grupo A (4,27) foi levemente inferior à do Grupo B (4,43).

No *box-plot* da Figura 1 é possível observar que o período formal de doutorado para 50% dos orientados de Computação foi de até 4 anos e 75% deles concluíram antes do quinto ano. O tempo mínimo do período de estudo foi de 0 anos e o máximo de 16. No histograma da mesma figura, observa-se que a moda do período foi de 4 anos. Os valores para os Grupos A e B foram similares.

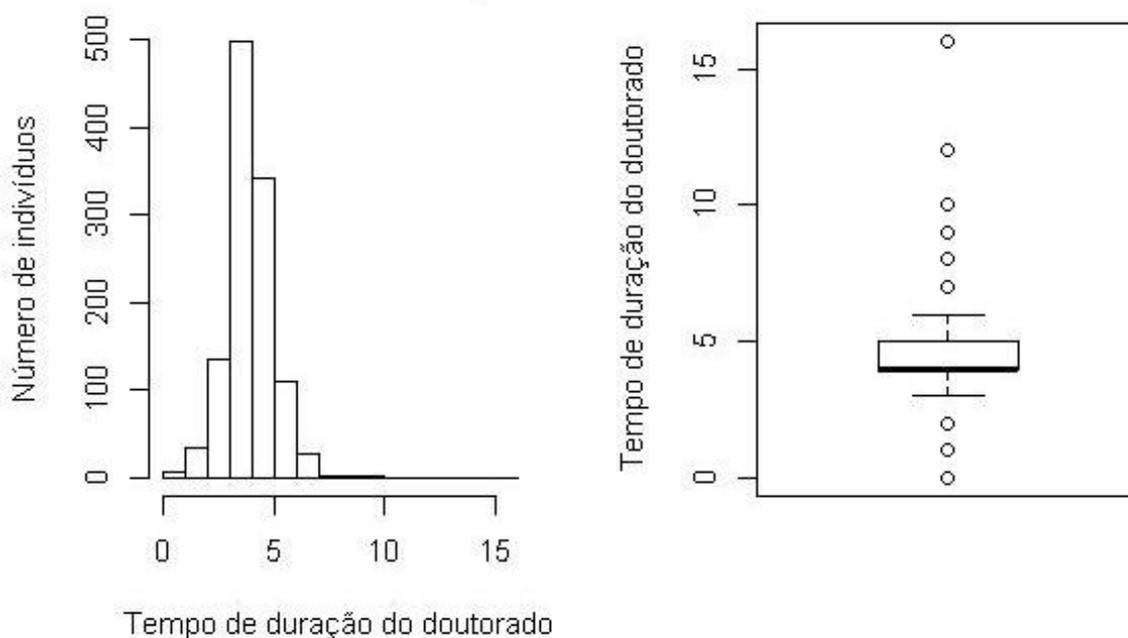


Figura 1. Quantidade de anos para a conclusão do doutorado

Fonte: elaborada pelos autores a partir de dados da pesquisa

### 3.1.2 Duração da colaboração

A duração do relacionamento foi calculada considerando também a presença ou não de publicações em conjunto com o orientador. No caso de não existir publicações em conjunto com o orientador, o tempo de relacionamento considerado foi a diferença entre o ano do fim e o ano do início do doutorado.

Pode-se observar, na Tabela 1, que a duração da colaboração para Computação estende-se em média por aproximadamente 5,58 anos, sendo, portanto, superior, em aproximadamente 1,5 anos à aparição da primeira publicação do par, e em aproximadamente 1 ano à aparição da primeira publicação do orientado sem a presença do orientador. Destaca-se ainda, na Tabela 1, que a duração média da colaboração para os indivíduos no Grupo A é superior em aproximadamente 2 anos à duração

para o Grupo B. Para o Grupo A o tempo médio de colaboração é superior em aproximadamente 3 anos ao fim do período formal de orientação e, no caso do Grupo B, não ultrapassa um ano.

Para Computação a porcentagem de pares para os quais não houve publicação conjunta entre orientado e orientador foi de 41,51%. Enquanto para o Grupo A foi de 18,4%, para o Grupo B foi de 44,68%.

### 3.1.3 A primeira publicação

A Figura 2 apresenta histogramas para analisar o tempo transcorrido até a primeira publicação, desde o fim do doutorado, segundo duas características: a presença ou não do orientador na publicação (as barras em amarelo indicam as publicações conjuntas com o orientador e as brancas indicam as publicações sem o orientador). É fácil observar que a maior proporção das primeiras publicações dos orientados esta concentrada no período de doutorado e que, passado este período, a proporção maior corresponde às publicações sem a presença do orientador.

Na Tabela 1, considerando o tempo para a primeira publicação desde a conclusão do doutorado, verificamos que a média para Computação é aproximadamente meio ano antes do fim do doutorado, com desvio padrão de 2,45 anos (não foram considerados os indivíduos que não publicaram com seus respectivos orientadores). Desta tabela também pode-se afirmar que, em média, a primeira publicação sem a presença do orientador aparece aproximadamente quando o orientado conclui o seu doutorado.

É importante notar que, para o Grupo A, o tempo médio até a primeira publicação sem o orientador é superior em aproximadamente 2 anos à média do tempo até a primeira publicação com o orientador. Enquanto, para o Grupo B, o tempo até a primeira publicação com ou sem o orientador são similares.

### 3.1.4 Número total de publicações em conjunto

Para Computação, a mediana desta distribuição foi de 1 artigo, a média foi 1,52 e o máximo de 24 artigos em conjunto. Em média, o número total de publicações em conjunto com o orientador do Grupo A é superior em aproximadamente 1,4 artigos ao Grupo B. Estes valores consideram todos os indivíduos de cada grupo, incluindo os que não publicaram.

O *box-plot* da Figura 3 mostra que 50% dos pares publicaram em conjunto até 1 artigo e 75% até 2 para Computação e para o Grupo B. No caso do Grupo A, 50% dos pares publicaram até 2 artigos e 75% até 4. No caso em que se consideraram apenas os indivíduos que publicaram pelo menos um artigo em conjunto, a mediana da distribuição de Computação foi de 2 artigos, a média foi 2,6. A média para o Grupo A foi 3,38 e para o Grupo B foi 2,45.

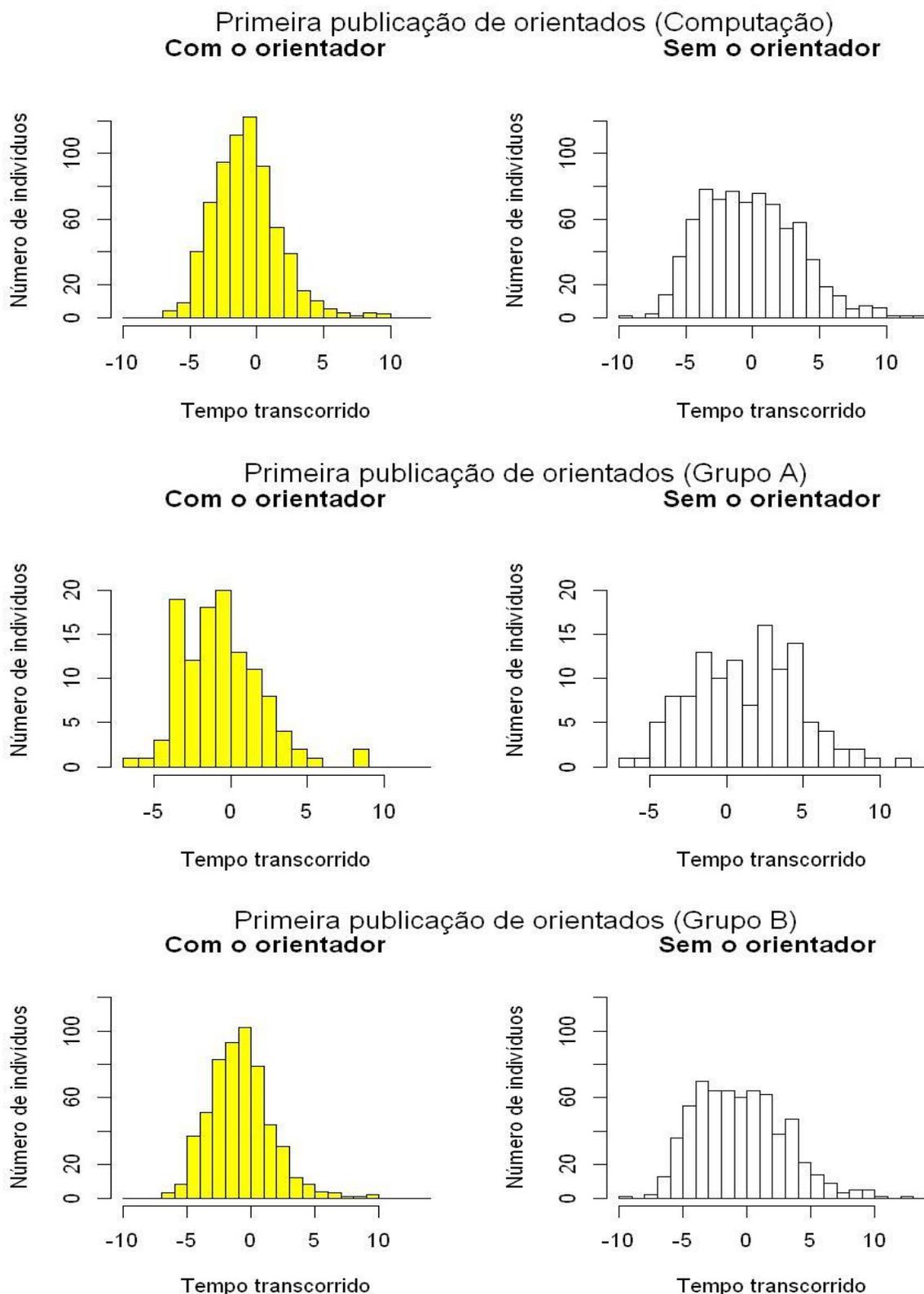


Figura 2. Histogramas do número de anos para a primeira publicação desde o fim do doutorado

Fonte: elaborada pelos autores a partir de dados da pesquisa

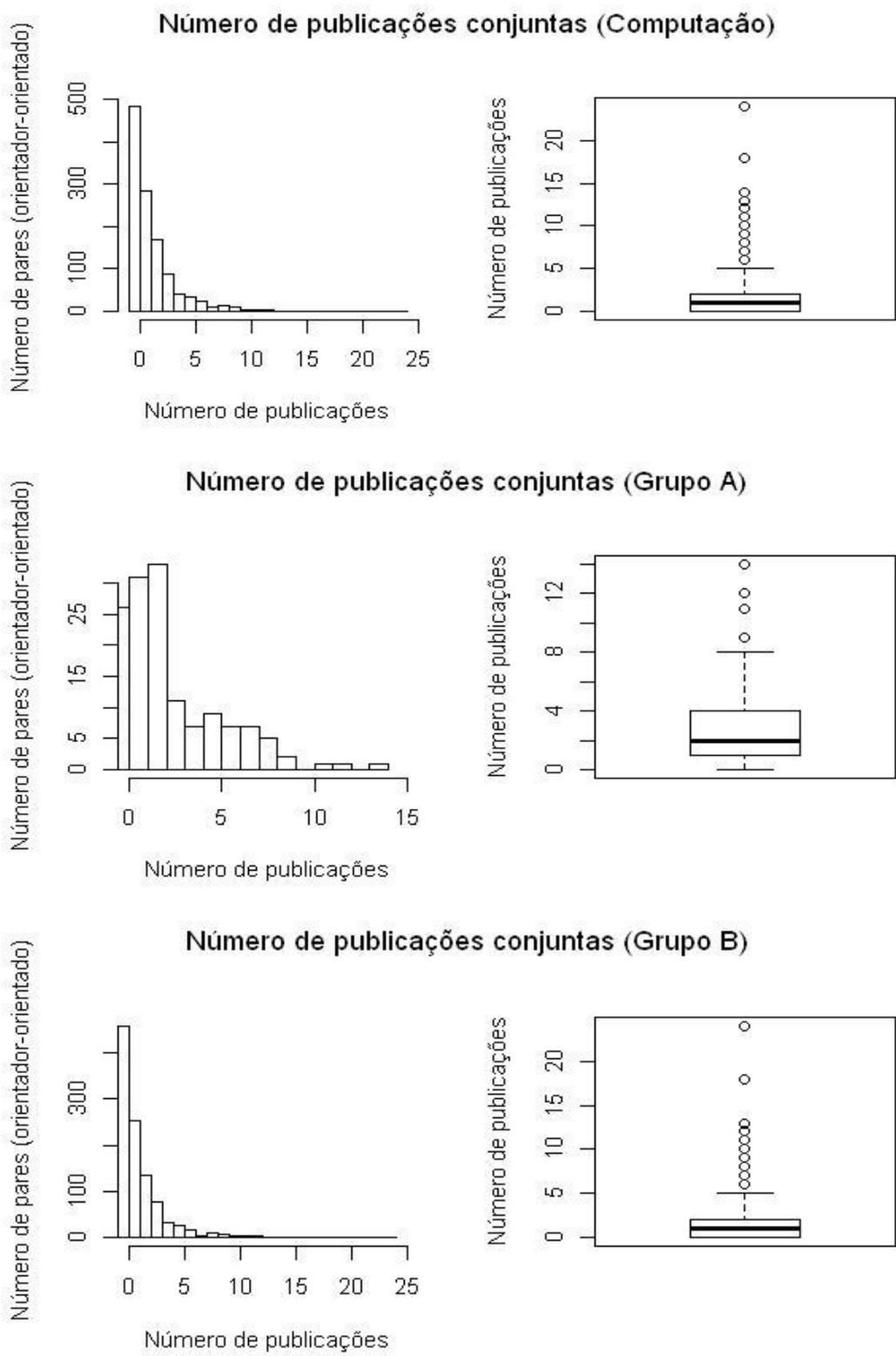


Figura 3. Número de publicações entre orientador e orientado  
 Fonte: elaborada pelos autores a partir de dados da pesquisa

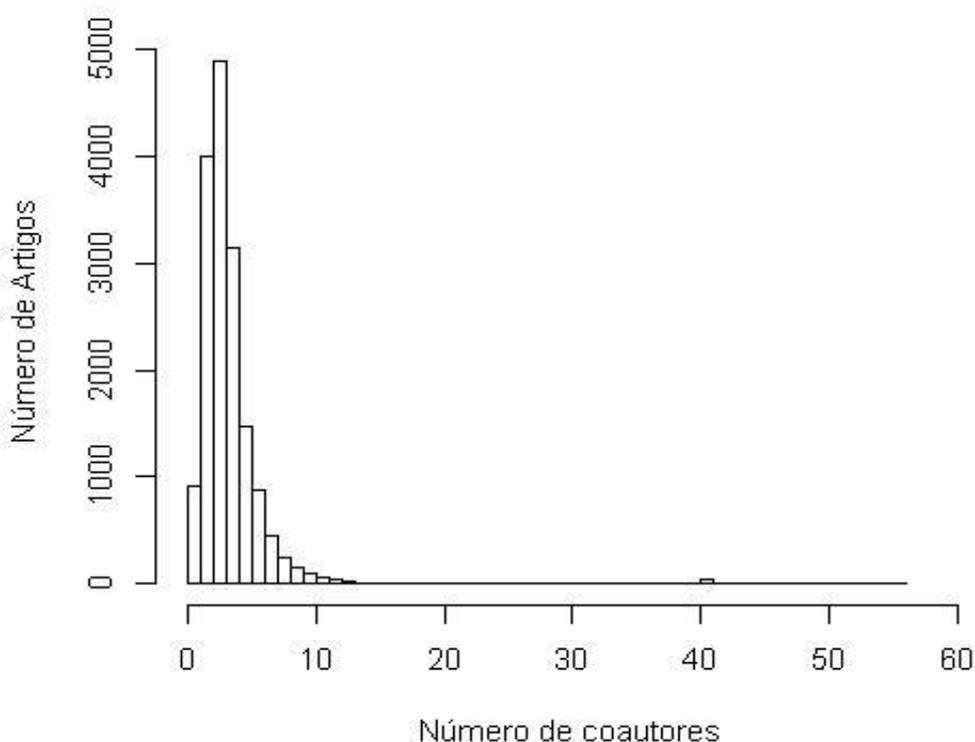


Figura 4. Histograma do número de coautorias para todas as publicações.

Fonte: elaborada pelos autores a partir de dados da pesquisa

### 3.1.5 Número de coautores

Foram encontrados 1.776 artigos escritos pelos pares orientador-orientado. Um vetor foi criado contendo o número de coautores para cada artigo, resultando em média, mediana e moda de 4,01, 3 e 3 autores, respectivamente. Outra informação interessante é a presença de 56 coautores em um artigo. Esta informação pode ser conferida no histograma apresentado na Figura 4 em que é possível observar uma leve assimetria à direita. Já na Tabela 1 observou-se que o número médio de coautores para as publicações conjuntas no Grupo A é levemente superior ao número médio de coautores para as publicações conjuntas no Grupo B (4,03 contra 3,86).

## 3.2 ANÁLISE DE CORRELAÇÃO

O coeficiente de correlação foi utilizado para calcular a relação entre a duração da colaboração (em anos) e o número total de publicações do orientado (Tabela 2). Para Computação o resultado foi de  $\rho = 0,32$ , com p-valor inferior a 0,01, mostrando a significância da relação entre estas variáveis. Para o Grupo A o coeficiente de correlação de Spearman foi aproximadamente duas vezes o do Grupo B.

Verificamos também que a correlação é menor para Computação e Grupo A quando considerado o total de publicações do orientador, ao invés do orientado. No entanto, o valor da correlação ainda é alto ( $\rho = 0.43$ ) para o Grupo A. Isto poderia ser um indicativo de que quanto maior o tempo da relação, maior será o número de publicações totais do orientador e do orientado, isto é, dado o perfil do orientado, ambos são beneficiados pela parceria.

Tabela 2. Correlação dos anos de relacionamento e número de artigos

	Computação		Grupo A		Grupo B	
	$\rho$ de Spearman	p-valor	$\rho$ de Spearman	p-valor	$\rho$ de Spearman	p-valor
Correlação entre anos de relacionamento do par e número total de artigos do orientado	0,32	$p^* < 0,01$	0,56	$p^* < 0,01$	0,26	$p^* < 0,01$
Correlação entre anos de relacionamento do par e número total de artigos do orientador	0,23	$p^* < 0,01$	0,43	$p^* < 0,01$	0,28	$p^* < 0,01$

Fonte: elaborada pelos autores a partir de dados da pesquisa

O coeficiente de correlação também foi utilizado para calcular a relação entre a duração da colaboração (em anos) e o *Author Rank* do orientado (Tabela 3). O *Author Rank* (que é uma modificação do *PageRank*) é uma medida que indica o grau de coautoria do pesquisador em relação aos outros pesquisadores considerados no conjunto em análise (LIU *et al.*, 2005).

Para o Grupo A o coeficiente de correlação de Spearman foi aproximadamente duas vezes o do Grupo B, com p-valor inferior a 0,01 ( $p^*$ ), mostrando a significância da relação entre estas variáveis.

Verificamos também que existe correlação entre o total de publicações do orientado e *Author Rank* do orientado e entre o total de publicações do orientador e o *Author Rank* do orientador (com p-valor inferior a 0,01) para ambos os grupos. Observe que o valor de  $\rho$  para os orientados é aproximadamente duas vezes o dos orientadores.

Note que a análise de correlação com o *Author Rank* na Tabela 2 permite diferenciar orientadores de orientados e ainda os grupos. Isto também é possível observar na diferença entre as médias e medianas do *Author Rank* para cada um dos grupos. Por exemplo, para o Grupo A, a média do *Author Rank* dos orientadores é 0,001876 enquanto a dos orientados é 0,00072.

Tabela 3. Correlação com o *Author Rank*

	Grupo A		Grupo B	
	$\rho$ de Spearman	p-valor	$\rho$ de Spearman	p-valor
Correlação anos de relacionamento do par e <i>Author Rank</i> do orientado	0,52	$p^* < 0,01$	0,21	$p^* < 0,01$
Correlação anos de relacionamento do par e número total de artigos do orientador	0,21	$p^* < 0,013$	0,04	$p < 0,2$
Correlação total de publicações do orientado e <i>Author Rank</i> do orientado	0,71	$p^* < 0,01$	0,61	$p^* < 0,01$
Correlação total de publicações do orientador e <i>Author Rank</i> do orientador	0,31	$p^* < 0,01$	0,24	$p^* < 0,01$
Correlação total de publicações conjuntas e <i>Author Rank</i> do orientado	0,62	$p^* < 0,01$	0,61	$p < 0,23$
Correlação total de publicações conjuntas e <i>Author Rank</i> do orientador	0,15	$p < 0,071$	0,14	$p^* < 0,01$

Fonte: elaborada pelos autores com base em dados da pesquisa

### 3.3 VISUALIZAÇÃO DAS REDES DE COAUTORIA E DE ORIENTAÇÃO

Com o objetivo de visualizar as redes são apresentadas duas figuras para representar tanto as relações de orientação como de coautoria. Na Figura 5, as arestas indicam a relação de orientação. Nas Figuras 5 e 6, os nós em azul representam os orientadores e os em verde os orientados. Nota-se que existem orientados que também orientaram pelo menos um indivíduo do conjunto.

Na Figura 6 as arestas indicam coautoria e apresentam espessuras diferentes, dependendo do número de trabalhos em conjunto. Os nós têm tamanhos diferentes, denotando um maior ou menor *Author Rank* do pesquisador representado pelo nó. A partir das Figuras 5 e 6, é possível observar a maior quantidade de relações de coautoria do que de orientação. Por exemplo, para o Grupo A nos grafos há 389 relações de coautoria e apenas 141 relações de orientação. Este fenômeno se justifica, em parte, pelo fato de uma orientação de doutorado gerar apenas uma relação de orientação, enquanto cada publicação pode envolver diversos pesquisadores da rede, gerando diversas relações de coautoria. Outras informações interessantes que podem ser observadas são a presença de alguns pesquisadores que são centralizadores de diferentes grupos de pesquisadores, alguns orientadores que são coautores de todas as publicações de seus orientandos, e algumas relações envolvendo três gerações de orientação.

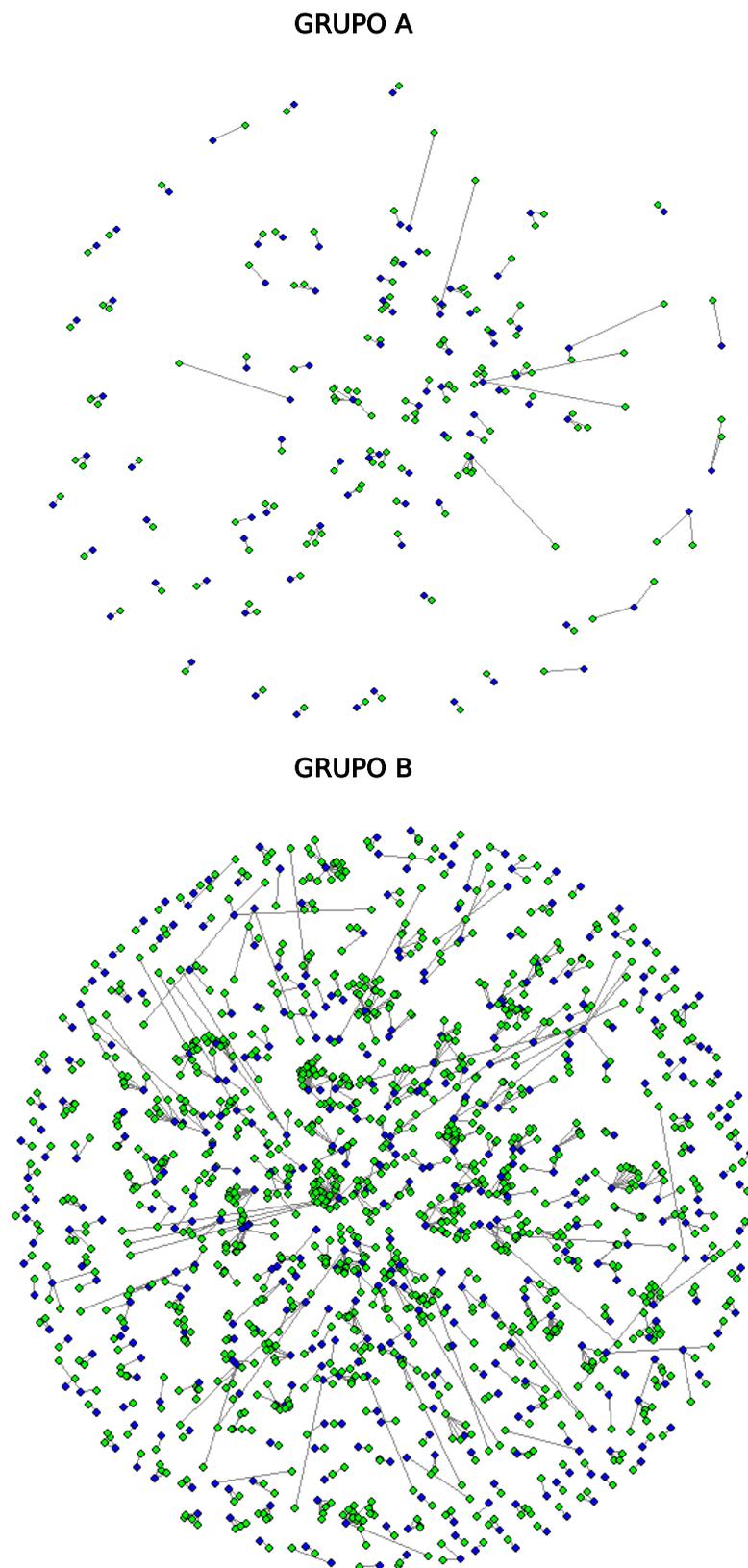


Figura 5. Rede de orientações acadêmicas de doutorado. Cada orientador é representado na cor azul. Cada orientado é representado na cor verde.

Fonte: elaborada pelos autores com base em dados da pesquisa

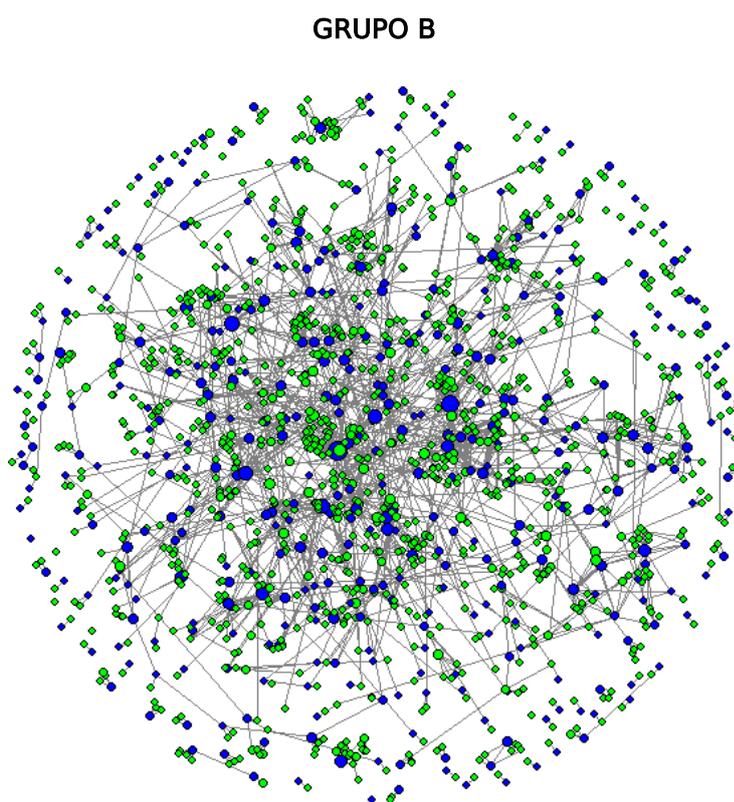
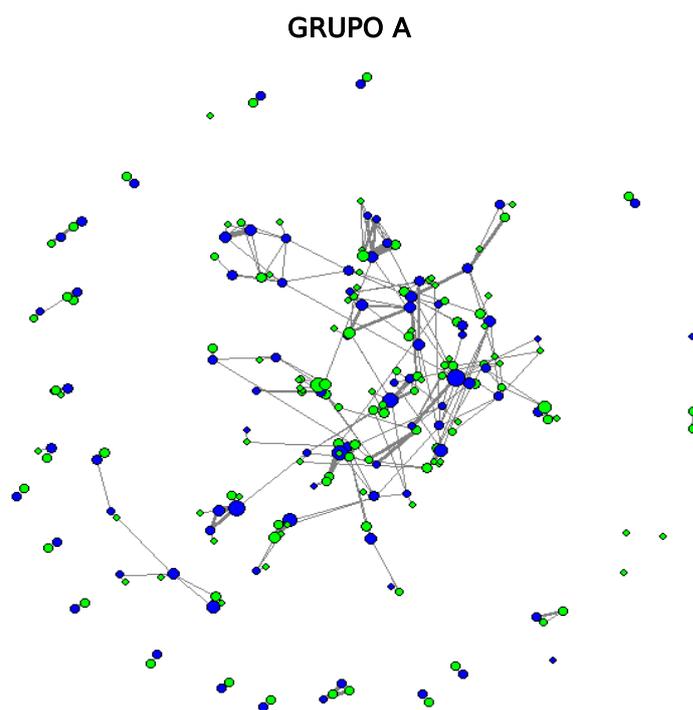


Figura 6. Rede de coautorias. Cada orientador é representado na cor azul. Cada orientado é representado na cor verde.

Fonte: elaborada pelos autores com base em dados da pesquisa

## 4 DISCUSSÃO

Nesta seção são discutidos os resultados correspondentes às questões levantadas na Seção 2.2.

Levando em consideração as mudanças no cenário da pós-graduação no Brasil, é possível ver que o período de doutoramento tem diminuído, provavelmente devido às facilidades trazidas pelo uso da Internet, a partir dos anos 90, e pelas maiores exigências das agências de fomento à pesquisa quanto ao período máximo no curso de pós-graduação. Para todos os grupos estudados verificou-se que as diferenças da duração desse período são mínimas e em média são próximas aos valores recomendados pelas agências.

Na amostra obtida dos currículos Lattes observou-se que a duração do tempo de relacionamento em média extrapola o período formal de orientação para todos os grupos estudados e que a parceria entre orientador e orientado não necessariamente gera algum artigo em periódico. Por exemplo, mais do que 40% dos pares em Computação e no Grupo B não publicaram artigos em periódico, com a ressalva de que estes indivíduos podem ter publicado outros artigos sem a presença do orientador ou outros tipos de trabalhos com o orientador. É sabido que orientadores e orientados em Ciência da Computação publicam preferencialmente em conferências, sendo aproximadamente 53% no triênio 2007-2009 (MENA-CHALCO, DIGIAMPIETRI, OLIVEIRA, 2012). Essa cultura está mudando, uma vez que as agências de fomento exigem que os pesquisadores escrevam artigos também em periódicos. Isto pode ser observado no fato de que a porcentagem dos pares que não publicaram em conjunto em periódicos no Grupo A é menor do que a metade (18,4%) da porcentagem dos outros grupos.

Ao se considerar apenas os pares que tiveram pelo menos uma publicação em conjunto, pode-se notar que a média de artigos em conjunto é levemente maior que 2,4 artigos para Computação e o Grupo B. Para o Grupo A, essa média é 3,38 artigos, o que também é um indicativo de que na academia é incentivada a independência do orientado em relação ao orientador, uma vez que a parceria para esse grupo gera apenas, em média, 1 artigo a mais quando comparado com o Grupo B.

Identificou-se também que a duração média do período de colaboração para os indivíduos do Grupo A é superior em aproximadamente três anos à data da primeira publicação conjunta e à data do fim do doutorado. Para os outros pares, esta diferença é de aproximadamente um ano. Notou-se a influência deste período maior de relacionamento ao se considerar que a média do número total de publicações dos orientados do Grupo A (6,8) é maior do que o dobro da média das publicações dos orientados do Grupo B (3,26) - considerados todos os pares, inclusive aqueles que não publicaram em conjunto. Esta influência também pode ser vista por meio do coeficiente de correlação de Spearman calculado para cada grupo.

Verificou-se ainda que as relações estabelecidas nas coautorias envolvem normalmente mais indivíduos do que apenas o orientador e seu orientado, o que é uma tendência na produção científica brasileira (DIGIAMPIETRI *et al.*, 2012).

Em resumo, é importante notar que o perfil do orientado pode variar segundo o grupo de atuação. Percebe-se que os indivíduos do Grupo A, quando comparados com os do Grupo B, apresentam: (i) menor proporção de pares que não publicaram em conjunto; (ii) maior tempo de relacionamento; (iii) maior concentração para a primeira publicação com o orientador do que sem o orientador; (iv) média de publicações em conjunto superior; (v) coeficiente de correlação de Spearman maior entre o número total de publicações (do orientador e do orientado) com o tempo de relacionamento; (vi) coeficiente de correlação de Spearman maior entre o *Author Rank* do orientado e anos de relacionamento e o total de publicações do orientado; e (vii) maior média de *Author Rank* dos orientados.

Existem duas limitações referentes à abordagem utilizada. A primeira está relacionada ao uso de dados diretamente extraídos dos currículos da Plataforma Lattes sem nenhuma validação adicional. Estes dados podem inserir ruídos na análise, referentes tanto a informações preenchidas incorretamente quanto devido a informações incompletas que ocorrem dada a falta de atualização dos currículos Lattes. A segunda limitação está relacionada com a determinação/escolha da medida do tempo de relacionamento que, neste trabalho, envolveu as datas da publicação de artigos em coautoria. É possível que o tempo de relacionamento se estenda além dessas datas. Acreditamos que, apesar destas limitações, a abordagem utilizada considera um volume bastante significativo de dados que permitiram obter resultados bastante interessantes.

## 5 CONSIDERAÇÕES FINAIS

O objeto deste estudo foram os pares formados pelos orientadores e orientados, em programas de doutorado da área de Ciência da Computação no Brasil com currículos cadastrados na plataforma Lattes. O objetivo principal foi investigar as características desses pares, em especial daqueles em que tanto o orientador quanto o orientado estavam cadastrados no corpo permanente de docentes de pós-graduação no triênio 2007-2009.

Uma suposição comum é que todos os alunos que estudam ou realizam um doutorado têm como principal objetivo trabalhar na área acadêmica. No entanto, especialmente em áreas tecnológicas, as ofertas do mercado podem atrair os pesquisadores, fazendo com que estes abandonem ou adiem os seus projetos de trabalhar como pesquisadores em universidades de prestígio. Ao investigar as características e o perfil dos orientados que seguem a carreira acadêmica, comparando-os com os indivíduos que, no período estudado, não estavam cadastrados em nenhum dos 45 programas de pós-graduação em Ciência da Computação do Brasil no período

2007-2009, deve estar claro que isto não implica que os orientados não cadastrados não estejam trabalhando em pesquisa, pois eles poderiam estar cadastrados em outros programas de pós-graduação ou também se encontrar trabalhando como pesquisadores em centros de investigação de empresas.

A realização deste trabalho permitiu a verificação da tendência internacional de mudanças dentro da área de Ciência da Computação, especialmente no que concerne aos tipos de documento utilizados para publicação, uma vez que se discute a vantagem de publicar em periódicos. Esta tendência tem sido também seguida pelos organismos de avaliação da pós-graduação no Brasil, que exigem que os pesquisadores nessa área também publiquem os seus trabalhos em periódicos internacionais reconhecidos.

Finalmente, é importante destacar que foi apresentada uma análise estatística da relação orientador-orientado que não permite observar a influência individual de um pesquisador na atuação acadêmica do outro, seja como número total de publicações bibliográficas ou como número de alunos formados pelo orientado. Nesse sentido, como trabalhos futuros, pretende-se dedicar esforços para estimar a provável influência entre orientador-orientado e vice-versa. Ainda, pretende-se expandir todas as análises para as outras áreas do conhecimento, como por exemplo, toda a Grande Área de Ciências Exatas e da Terra.

## AGRADECIMENTOS

O trabalho apresentado neste artigo foi parcialmente financiado pela FAPESP (Projeto Jovem Pesquisador processo 2009/10413-5) e pelo CNPq (Bolsa Produtividade em Pesquisa processo 304937/2010-0).

## REFERÊNCIAS

AMARAL, L. (JC e-mail, n. 4075, 16 ago. 2010). Termina processo de Avaliação Trienal 2007-2009 da Capes. Disponível em: <http://www.jornaldaciencia.org.br/Detail.jsp?id=72830>. Acessado em 10 set. 2011.

ANWAR, M. From doctoral dissertation to publication. A study of 1995 American graduates in library and information science. *Journal of Librarian and Information Science*, v. 36, n. 4, p.151–157, 2004.

ARRIOLA-QUIROZ, I.; CURIOSO, W.; CRUZ-ENCARNACION, M.; GAYOSO, O. Characteristics and publication pattern of theses from a Peruvian medical school. *Health Information and Libraries Journal*, v. 27, p. 148–154, 2010. <http://dx.doi.org/10.1111/j.1471-1842.2010.00878.x>

DIGIAMPIETRI, L.; MENA-CHALCO, J. P.; PÉREZ-ALCÁZAR, J. J.; TUESTA, E. F.; DELGADO, K.; MUGNAINI, R.; SILVA, G. Minerando e Caracterizando Dados de Currículos Lattes. *I Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2012)*, 2012.

FRKOVIC, V.; SKENDER, T.; DOJCINOVIC, B. Publishing scientific papers based on master's and Ph.D. theses from a small community: case study Croatian medical schools. *Croatian Medical Journal*, v. 44, n. 1, p.107–111, 2003.

GOLDE, C. M. Should I stay or should I go? Student descriptions of the doctoral attrition process. *Review of Higher Education*, v. 23, n. 2, p. 199–227, 2000. <http://dx.doi.org/10.1353/rhe.2000.0004>

KAM, B. H. Style and quality in research supervision: the supervisor dependency factor. *Higher Education*, v. 34, n. 1, p. 81–103, 1997. <http://dx.doi.org/10.1023/A:1002946922952>

LARIVIERE, V. On the shoulders of students? The contribution of Ph.D. students to the advancement of knowledge. *11th International Conference on STI "Creating Value for Users"*. Leiden, Netherlands, 2010.

LEE, W. M. Publication trends of doctoral students in three fields from 1965-1995. *Journal of the American Society for Information Science*, v. 51, n. 2, p. 139–144, 2000. [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:2<139::AID-ASI5>3.0.CO;2-1](http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:2<139::AID-ASI5>3.0.CO;2-1)

LIU, X.; BOLLEN, J.; NELSON, M. L.; DE SOMPEL, H. V. Co-authorship networks in the digital library research community. *Inf. Process. Manage*, v. 41, n. 6, p. 1462–1480, 2005. <http://dx.doi.org/10.1016/j.ipm.2005.03.012>

MAINHARD, T.; VAN DER RIJST, R.; VAN TARTWIJK, J.; WUBBELS, T. A model for the supervisor doctoral student relationship. *Higher Education*, v. 58, n. 3, p. 359–373, 2009. <http://dx.doi.org/10.1007/s10734-009-9199-8>

MALLETTE, L. Publishing rates of graduate education Ph.D. and Ed.D. students: a longitudinal study of University of California schools. Doctoral dissertation, Pepperdine University. Retrieved from ProQuest Dissertations & Theses database. (UMI No. 3239922), 2006.

MARSH, H. W.; ROWE, K. J.; MARTIN, A. Ph.D. students? Evaluations of research supervision. *Higher Education*, v. 73, n. 3, p. 313–348, 2002. <http://dx.doi.org/10.1353/jhe.2002.0028>

McALPINE, L.; NORTON, J. Reframing our approach to doctoral programs: an integrative framework for action and research. *Higher Education Research and Development*, v. 25, n. 1, p. 3–17, 2006. <http://dx.doi.org/10.1080/07294360500453012>

MENA-CHALCO, J. P., CESAR JR, R. M. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, v. 15, p. 31-39, 2009. <http://dx.doi.org/10.1007/BF03194511>

MENA-CHALCO, J. P.; CESAR-JR., R. M. Towards automatic discovery of coauthorship networks in the Brazilian academic areas. *In IEEE Seventh International Conference on e-Science Workshops 2011 (eScienceW)*, p. 53–60, 2011.

MENA-CHALCO, J. P.; DIGIAMPIETRI, L. A.; OLIVEIRA, L. B. Perfil de produção acadêmica dos programas brasileiros de pós-graduação em ciência da computação nos triênios 2004-2006 e 2007-2009. *Em Questão*, v. 18, n. 3, p. 215-229, 2012.

MUGNAINI, R.; IGAMI, M. P. Z.; BRESSIANI, J. C. Productivity and doctoral research in a Brazilian Nuclear Research institution: validating co-word analysis technique. *Proceedings of the ISSI 2011 Conference*. Durban: University of Zululand Reprographic and Printing Centre, 2011., v.II, p. 1037-1039, 2011.

MYERS, J. L.; WELL, A. D. Research design and statistical analysis. 2. edition. Lawrence Erlbaum, 2003.

POBLACIÓN, D.; MUGNAINI, R.; RAMOS, L. Redes sociais e colaborativas em informação científica. São Paulo : Angellara Editoras, 2009.

RAMOS, P.; FURTADO, E.; CARVALHO, E.; CAMPOS, M.; CORTES DE SOUZA, D.; ALMEIDA, L.; DE ARAUJO, C. Dissertações e teses de pós-graduação geram publicação de artigos científicos? Análise baseada em três programas da área de Educação Física. *Brazilian Journal of Biomotricity*, v. 3, n. 4, p. 315-324, 2009.

SACARDO, M.; HAYASHI, M. C. Balanço bibliométrico da produção científica em Educação Física e Educação Especial oriunda de teses e dissertações. *Revista Brasileira de Pós-Graduação*, v. 8, n. 15, p. 111-135, 2011.

SALMI, L.; GANA, S.; MOUILLET, E. Publication pattern of medical theses, France 1993-98. *Medical Education*, v. 35, n. 1, p. 18-21, 2001. <http://dx.doi.org/10.1046/j.1365-2923.2001.00768.x>

TUESTA, E. F.; DELGADO, K.; DIGIAMPIETRI, L.; MUGNAINI, R.; MENA-CHALCO, J. P.; PÉREZ-ALCÁZAR, J. J. Análise temporal da relação orientador-orientado: um estudo de caso sobre a produtividade dos pesquisadores doutores da área de Ciência da Computação. *I Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2012)*, 2012.

WANG, C.; HAN, J.; JIA, Y.; TANG, J.; ZHANG, D.; YU, Y.; GUO, J. Mining advisor-advisee relationships from research publication networks. *In Knowledge Data Discovery (KDD)*, p. 203-212, 2010. <http://dx.doi.org/10.1145/1835804.1835833>

WANG, C.; HAN, J.; LI, Q.; LI, X.; LIN, W.-P.; JI, H. Learning hierarchical relationships among partially ordered objects with heterogeneous attributes and links. *In Proceedings of 2012 SIAM International Conference on Data Mining*, 2012. <http://dx.doi.org/10.1137/1.9781611972825.45>

WU, T.; CHEN, Y.; HAN, J. Association mining in large databases: a reexamination of its measures. *In PKDD*, p. 621-628, 2007.

WU, T.; CHEN, Y.; HAN, J. Re-examination of interestingness measures in pattern mining: a unified framework. *Data Min. Knowl. Discov.*, v. 21, n. 3, p. 371-397, 2010. <http://dx.doi.org/10.1007/s10618-009-0161-2>