

Análise temporal da relação orientador-orientado: um estudo de caso sobre a produtividade dos pesquisadores doutores da área de Ciência da Computação

Esteban F. Tuesta¹, Karina V. Delgado¹, Luciano A. Digiampietri¹,
José J. P. Alcázar¹, Rogério Mugnaini¹, Jesús P. Mena-Chalco²

¹Escola de Artes, Ciências e Humanidades
Universidade de São Paulo (EACH-USP) – São Paulo – SP – Brazil

²Centro de Matemática, Computação e Cognição
Universidade Federal do ABC (UFABC) – São Paulo, SP – Brazil

{tuesta, kvd, digiampietri, jperez, mugnaini}@usp.br,

jesus.mena@ufabc.edu.br

Abstract. *This paper presents a study of the academic relationship between advisor and advisee for a group of Ph.D. researchers of the Computer Science Area with curricula registered in the Lattes Platform. We analyze the main characteristics of the group and their coauthor relationships. In this analysis we observe that the duration of the co-authoring between advisor and advisee can extrapolate the formal period of supervision, also, we show that the duration of the collaboration, as coauthors, is correlated with the number of articles published by the advisee.*

Resumo. *Este artigo apresenta um estudo da relação acadêmica temporal entre orientador e orientado, para um conjunto de pesquisadores doutores da área de Ciência da Computação com currículos cadastrados na plataforma Lattes. Analisamos as principais características do grupo e suas relações de coautoria. Observamos que o tempo de duração da coautoria, que se estabelece entre orientador e orientado, pode extrapolar o período formal de orientação, e, além disso, mostramos que a duração do tempo de colaboração, na forma de coautoria, correlaciona-se com o número de artigos em revistas, publicadas pelo orientado.*

1. Introdução e Motivação

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 revistas 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 culmina na publicação de artigos. Dados sobre a formação de doutores no Brasil mostram que desde o ano 1981 até 2009 se formaram 96.782 doutores¹.

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

Um estudo recente sobre a produtividade dos alunos do Programa de Pós-Graduação do Instituto de Pesquisas Energéticas e Nucleares, em seus mais de 30 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 nos últimos 10 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 vem 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 and 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 revistas. 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: 5 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 and Norton 2006] também indicam a importância dessa relação.

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 consiste em estudar a 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. Esse conjunto de pesquisadores encontra-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, o tempo de duração da coautoria estabelecida entre orientador e orientado (i) pode extrapolar o período formal de orientação, e (ii) correlaciona-se com o número de produções bibliográficas do aluno orientado.

Este artigo está organizado da seguinte forma: na Seção 2, apresentamos as noções das redes de coautorias, na Seção 3, descrevemos o processo de seleção da amostra e as

análises principais realizadas. Finalmente, as considerações e resultados finais do trabalho são apresentadas na na Seção 4.

2. Redes de coautorias

Todo conglomerado social é construído ao redor de relacionamentos, a atuação de um indivíduo não pode ser totalmente entendida a menos que a relacionemos às ações de outros com quem esteja conectado através de ligações sociais. Esse conglomerado formado pelos indivíduos e suas ligações é chamado de uma rede social, cujo estudo, visualização e caracterização vêm sendo denominados "análise de redes sociais". Diante desta demanda de pesquisa, estudos vêm sendo desenvolvidos na área de ciências exatas, mais especificamente no contexto da teoria de redes e teoria dos grafos [Wasserman and Faust 1994]. Nesses estudos são propostas diversas métricas topológicas para analisar diferentes características de uma rede social.

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 and Cesar-Jr. 2011], e avaliar o status dos pesquisadores de acordo com algumas métricas [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, em que existe um compromisso formal entre os indivíduos de produzirem algum produto de natureza científica.

3. Amostra da comunidade de pesquisadores em ciências da computação

3.1. Procedimento para obtenção da amostra

Os dados analisados neste artigo foram extraídos dos currículos da Plataforma Lattes². Segundo informações disponibilizadas pelo CNPq³ se formaram 1746 doutores em Ciência da Computação entre os anos 1981 e 2009. A identificação dos 889 currículos pertencentes aos docentes do corpo permanente dos 45 programas de pós-graduação acadêmica (que possuem doutorado ou mestrado acadêmico) em computação presentes na última avaliação, triênio 2007-2009, da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) foi feita através de um processo semiautomático considerando os *Cader-nos de Indicadores*.

Após a identificação dos currículos, estes foram baixados através do scriptLattes [Mena-Chalco and Cesar-Jr. 2009], tabulados e salvos em um banco de dados por um *parser* que discriminou cada um dos itens do currículo. Para este artigo, cada currículo foi dividido nas formações do pesquisador, orientações e publicações em periódicos.

As relações de orientação foram estabelecidas através de uma verificação dupla entre os currículos analisados. Foram consideradas orientações apenas as relações 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.

²<http://lattes.cnpq.br/>

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

Utilizando estes critérios, dos 889 currículos analisados, 239 possuem ao menos uma relação de orientação com outro(s) integrante(s) do conjunto. Estes currículos foram selecionados e as relações de coautoria de artigos publicados em periódicos foram examinadas. Para a identificação das relações de coautoria foi utilizado um algoritmo de casamento inexato de *strings* para comparar os títulos das publicações e a lista de coautores.

3.2. Descrição da amostra

Os dados disponíveis de cada um dos 239 pesquisadores selecionados são as publicações por ano em periódicos, sexo, ano de início de doutorado, ano de conclusão de doutorado, ano da última atualização do currículo Lattes, número de coautores de cada publicação. Além disso, para cada par orientador-orientado, dispomos das publicações em conjunto assim como o número de coautores destas publicações. Tanto os anos de todas as publicações como os anos de início de doutorado e conclusão de doutorado, não são superiores a 2010 nem inferiores a 1981.

No conjunto estudado, encontramos a presença de 146 orientados e 98 orientadores, 5 indivíduos do primeiro grupo, também participaram do segundo (orientaram pelo menos um indivíduo do conjunto). Quanto ao gênero, na amostra há 172 homens (aproximadamente 72%) e 67 mulheres (aproximadamente 28%). Foram considerados para análise, apenas os artigos em periódicos, dos quais 367 artigos corresponderam a trabalhos em conjunto (orientador e orientado), 2.466 artigos escritos pelos orientadores e 1.198 pelos orientados. É importante esclarecer que, na contagem dos artigos publicados individualmente pelos orientadores e orientados estão considerados os 367 artigos publicados em conjunto.

Para análise dos dados, foi utilizado o pacote estatístico do R-project versão 2.14.2⁴.

3.3. Análise descritiva da rede

A análise descritiva revela alguns resultados importantes, como por exemplo, a Figura 1, que consiste de um histograma do número de coautores para todas as publicações dos indivíduos da amostra, apresenta a moda de 3 coautores por artigo, e leve assimetria à direita.

Já a Tabela 1 permite comparar o número de coautores para o total de publicações com aquelas que envolvem apenas orientador e orientado. Podemos ver que enquanto a moda do número de coautores é exatamente a mesma (3), a média apresenta-se superior quando orientador e orientado estão presentes (4,11 contra 3,55 para o total de publicações). A razão principal desta diferença poderia residir no fato de que, para as coautorias orientador-orientado, o número mínimo de coautores é de 2 indivíduos, já na amostra total observamos que há 147 artigos publicados por um único autor. Outra informação interessante é a presença de 31 coautores em um artigo.

Observamos na amostra obtida dos currículos Lattes, que a parceria entre orientador e orientado nem sempre gera algum artigo em revista (Figura 2), desde que 32 doutores não publicaram com seus respectivos orientadores (aproximadamente 22% da

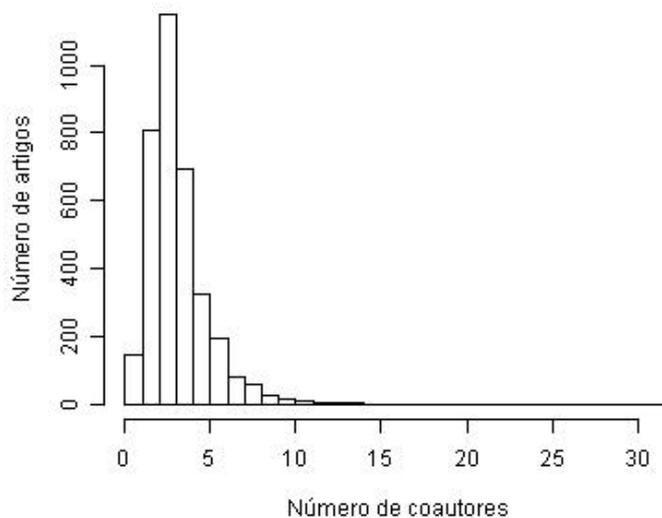
⁴<http://www.r-project.org/>

Tabela 1. Medidas descritivas

Item	Moda	Média	Desvio Padrão
Número de coautores de todas as publicações	3	3,55	1,79
Número de coautores em publicações orientador-orientado	3	4,11	1,91
Número de publicações orientador-orientado	1	2,48	2,63
Duração do doutorado	4	4,30	1,16
Duração do relacionamento orientador-orientado		7,29	4,04
Tempo transcorrido até a primeira publicação orientador-orientado desde o início do doutorado		4,53	3,13
Tempo transcorrido até a primeira publicação orientador-orientado desde o fim do doutorado		0,42	3,04
Tempo transcorrido até a primeira publicação sem a presença do orientador desde o fim do doutorado		1,32	3,46
Número de publicações dos orientadores		25,16	18,62
Número de publicações dos orientados		8,21	7,88

amostra) 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. A moda desta distribuição foi de 1 artigo, a média foi 2,48 e o máximo de 14 artigos em conjunto. O box-plot da Figura 2 mostra que 50% dos pares publicaram em conjunto até 2 artigos e 75% até 4.

A Figura 3, mostra dois 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 con-

**Figura 1. Histograma do número de coautorias para todas as publicações.**

juntas e as brancas indicam as publicações sem o orientador). É fácil observar que a maior proporção das primeiras publicações dos orientados estão concentradas no período de doutorado e que passado este período, a proporção maior corresponde às publicações sem a presença do orientador.

Quanto ao tempo que um aluno demora a concluir seu doutorado, observamos na Tabela 1 e na Figura 4 que a moda foi de 4 anos e a média de 4,3 anos. Esta mesma análise mostra que 50% dos orientados concluíram os seus estudos em até 4 anos e 75% antes do quinto ano (a contar desde o início do doutorado). O tempo mínimo do período de estudo foi de 2 anos e o máximo de 12.

Na Tabela 1, considerando o tempo para a primeira publicação desde a conclusão

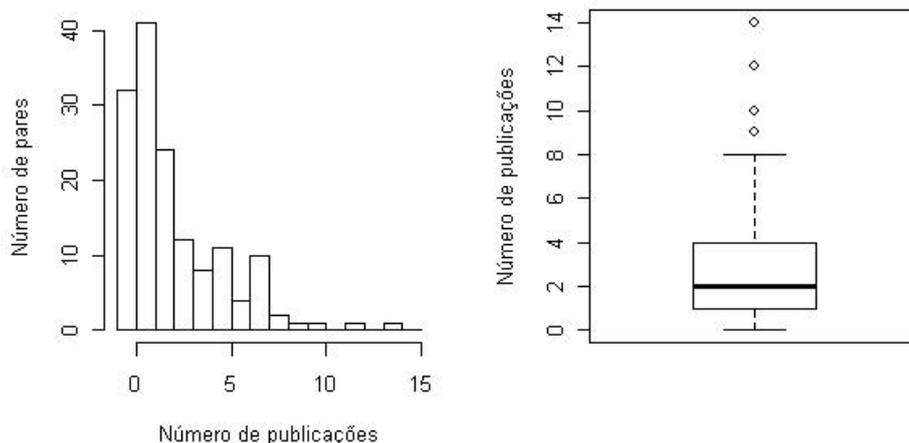


Figura 2. Número de publicações entre orientador e orientado.

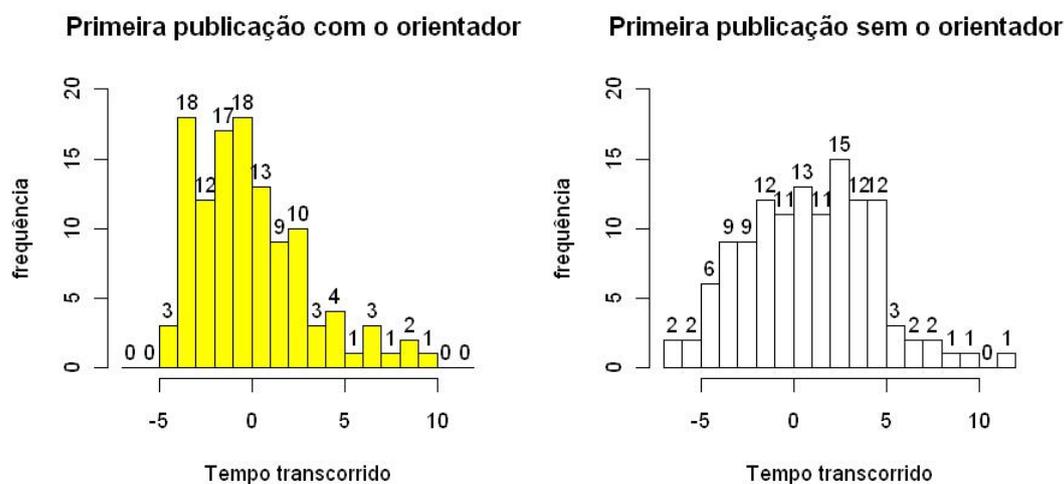


Figura 3. Histogramas do número de anos para a primeira publicação desde o fim do doutorado.

do doutorado, verificamos que a média é de 0,42 anos, com desvio padrão de 3,04 anos (aqui não estamos considerando os indivíduos que não publicaram com seus respectivos orientadores). Isto implica que, na maioria das orientações, a primeira publicação conjunta aparece menos de um ano após a conclusão do doutorado. Desta tabela também podemos ver que, em média, a primeira publicação sem a presença do orientador aparece aproximadamente um ano após a primeira publicação do par.

Podemos também observar na Tabela 1 que a duração da colaboração (definida neste trabalho 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, a definimos como a diferença entre o ano de conclusão e o ano de início do doutorado), estende-se em média pelo período de aproximadamente 7,29 anos, sendo portanto superior, em média, em aproximadamente 3 anos à aparição da primeira publicação do par, e em 2 anos à aparição da primeira publicação do orientado sem a presença do orientador.

O número médio de publicações dos orientados é de 8,21, com desvio padrão de 7,88, e o número médio de publicações do orientador é de 25,16 com desvio padrão de 18,62.

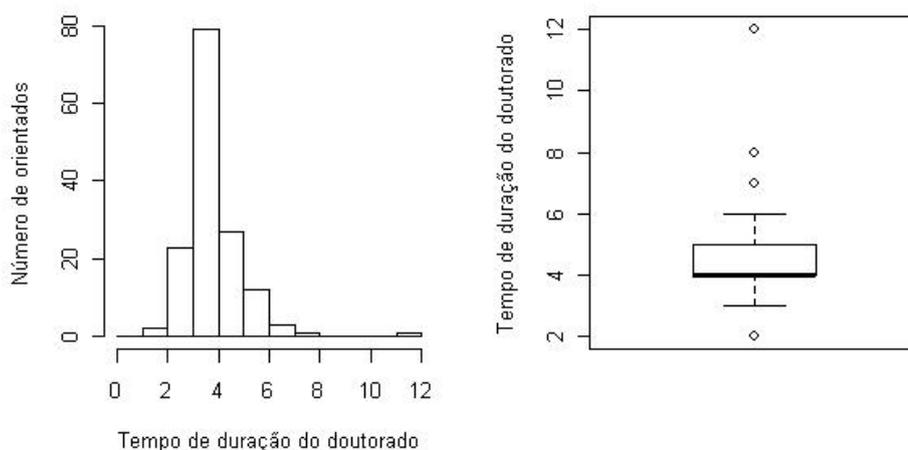


Figura 4. Quantidade de anos para a conclusão do doutorado.

3.4. Visualização da rede

Com o objetivo de visualizar as redes construímos duas figuras para representar as relações de orientação, coautoria e *Author Rank*. 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]. Nas Figuras 5 e 6 os nós em azul representam os orientadores, em amarelo os orientados e em verde os orientados que também orientaram pelo menos um indivíduo do conjunto.

Na Figura 6 as arestas cor cinza indicam coautoria e apresentam espessuras diferentes, dependendo do número de trabalhos em conjunto. Na Figura 5 as arestas vermelhas indicam a relação de orientação. Os nós têm tamanhos diferentes, denotando o maior ou

menor *Author Rank* do pesquisador representado pelo nó. Ordenando os valores do *Author Rank* observamos que o maior valor é de 17,47.

Das Figuras 5 e 6 é possível observar a maior quantidade de relações de coautoria do que de orientação. De fato, nos grafos dessas figuras há 319 relações de coautoria e apenas 148 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: o primeiro orientador em azul, seu orientando em verde que, por sua vez, já orientou um ou mais indivíduos da rede.

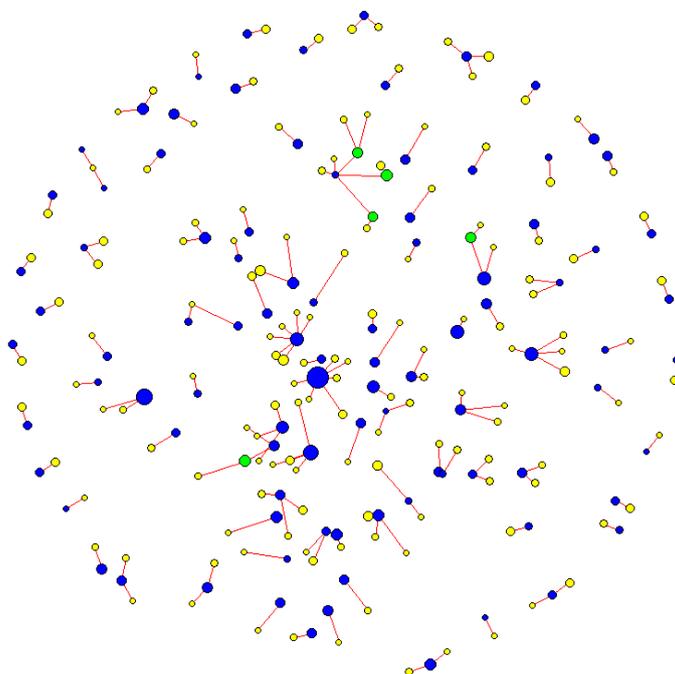


Figura 5. Rede de orientações acadêmicas de doutorado.

3.5. Análise de correlação

Na estatística, uma medida não paramétrica usada para mensurar a força de associação entre duas variáveis é o coeficiente de correlação de Spearman (ρ de Spearman), que se diferencia do coeficiente de correlação de Pearson por não exigir normalidade das variáveis, nem de linearidade da associação entre as mesmas. Quando existe a perfeita relação, o valor de ρ será 1 ou -1, conforme a associação for positiva ou negativa, e um valor de ρ igual a zero ou próximo de zero indica a falta de associação. O teste de hipótese sobre o coeficiente, usa procedimentos assintóticos [Myers and Well 2003].

Este coeficiente de correlação foi utilizado para calcular a relação entre Anos de Relacionamento (tempo desde o início do doutorado até a última publicação), e o número

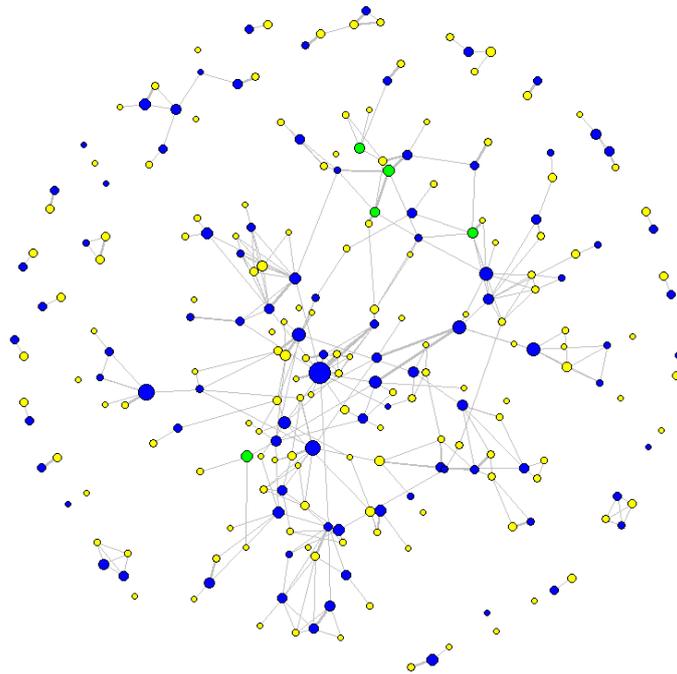


Figura 6. Rede de coautorias.

total de publicações do orientado. O resultado foi de $\rho = 0,51$, com p-valor inferior a 0,001, mostrando a significância da relação entre estas variáveis. Isto é um indicativo claro de que quanto maior o tempo da relação, maior será o número de publicações totais do orientado.

A segunda relação encontrada foi entre as variáveis ano de início do doutorado e duração do doutorado. O coeficiente de correlação de Spearman foi de $-0,22$ ($\rho = -0,22$) e p-valor inferior a 0,009, mostrando que ao longo dos anos o período do doutorado tem diminuído paulatinamente.

4. Conclusões e trabalhos futuros

Neste trabalho investigamos características da relação entre pares orientador-orientado que pertencem a programas de pós-graduação em Ciências da Computação no Brasil no triênio 2007-2009. Identificamos que a duração média do período de colaboração é superior em aproximadamente 3 anos à data da primeira publicação e à data do fim do doutorado. 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 maiores 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.

Outro resultado importante encontrado a partir da análise de correlação de Spearman indica a relação positiva ($\rho = 0,51$) e significativa ($p < 0,001$) entre o período de duração do relacionamento e o número total de publicações do orientado, o que pode indicar que o orientado é beneficiado no sentido de ter maior número de publicações,

quando aumenta o período de relacionamento entre ele e o orientador. Verificamos 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].

Propomos como trabalhos futuros o uso de outras variáveis para a caracterização da relação entre pares, entre elas, as notas das avaliações trienais de cada programa de pós-graduação em computação feita pela CAPES, o sexo e as respectivas idades dos indivíduos.

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/Detailhe.jsp?id=72830>. Acesso em 10 set. 2011.
- Anwar, M. (2004). From doctoral dissertation to publication. a study of 1995 American graduates in library and information science. *Journal of Librarian and Information Science*, 36 (4):151–157.
- Arriola-Quiroz, I. and Curioso, W., Cruz-Encarnacion, M., and Gayoso, O. (2010). Characteristics and publication pattern of theses from a Peruvian medical school. *Health Information and Libraries Journal*, 27:148–154.
- Digiampietri, L. A., Mena-Chalco, J. P., Pérez-Alcázar, J. J., Tuesta, E. F., Delgado, K. V., Mugnaini, R., and Silva, G. S. (2012). Minerando e caracterizando dados de currículos lattes. In *Proceedings of BraSNAM 2012*.
- Frkovic, V., Skender, T., and Dojcinovic, B. (2003). Publishing scientific papers based on master's and Ph.D. theses from a small community: case study Croatian medical schools. *Croatian Medical Journal*, 44(1):107–111.
- Golde, C. M. (2000). Should I stay or should I go? student descriptions of the doctoral attrition process. *Review of Higher Education*, 23(2):199–227.
- Kam, B. H. (1997). Style and quality in research supervision: the supervisor dependency factor. *Higher Education*, 34(1):81–103.
- Lariviere, V. (2010). 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.
- Lee, W. M. (2000). Publication trends of doctoral students in three fields from 1965-1995. *Journal of the American Society for Information Science*, 51(2):139–144.
- Liu, X., Bollen, J., Nelson, M. L., and de Sompel, H. V. (2005). Co-authorship networks in the digital library research community. *Inf. Process. Manage.*, 41(6):1462–1480.
- Mainhard, T., van der Rijst, R., van Tartwijk, J., and Wubbels, T. (2009). A model for the supervisor doctoral student relationship. *Higher Education*, 58(3):359–373.
- Mallete, L. (2006). Publishing rates of graduates 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).

- Marsh, H. W., Rowe, K. J., and Martin, A. (2002). Ph.D. students' evaluations of research supervision. *Higher Education*, 73(3):313–348.
- McAlpine, L. and Norton, J. (2006). Reframing our approach to doctoral programs: an integrative framework for action and research. *Higher Education Research and Development*, 25(1):3–17.
- Mena-Chalco, J. P. and Cesar-Jr., R. M. (2009). scriptLattes: An open-source knowledge extraction system from the lattes platform. *Journal of the Brazilian Computer Society*, 15(4):31–39.
- Mena-Chalco, J. P. and Cesar-Jr., R. M. (2011). Towards automatic discovery of co-authorship networks in the Brazilian academic areas. In *IEEE Seventh International Conference on e-Science Workshops 2011 (eScienceW)*, pages 53–60. IEEE. (Workshop On Measuring the Impact of e-Science Research – MeSR 2011).
- Mugnaini, R., Igami, M. P. Z., and Bressiani, J. C. (2011). 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:1037–1039.
- Myers, J. L. and Well, A. D. (2003). *Research Design and Statistical Analysis*. Lawrence Erlbaum, 2nd edition.
- Población, D., Mugnaini, R., and Ramos, L. (2009). *Redes sociais e colaborativas em informação científica*. Angellara Editoras, São Paulo, 1st edition.
- Ramos, P., Furtado, E., Carvalho, E., Campos, M., Cortes de Souza, D., Almeida, L., and de Araujo, C. (2009). Dissertações e Teses de Pós Graduação geram publicação de artigos científicos? análise baseada em 3 programas da área de educação física. *Brazilian Journal Biomotricity*, 3(4):315–324.
- Sacardo, M. and Hayashi, M. C. (2011). Balanço bibliométrico da produção científica em Educação Física e Educação Especial oriunda de teses e dissertações. *RBPG Revista Brasileira de Pós-Graduação*, 8(15):111–135.
- Salmi, L., Gana, S., and Mouillet, E. (2001). Publication pattern of medical theses, France 1993-98. *Medical Education*, 35(1):18–21.
- Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., and Guo, J. (2010). Mining advisor-advisee relationships from research publication networks. In *Knowledge Data Discovery (KDD)*, pages 203–212.
- Wang, C., Han, J., Li, Q., Li, X., Lin, W.-P., and Ji, H. (2012). Learning hierarchical relationships among partially ordered objects with heterogeneous attributes and links. In *Proceedings of 2012 SIAM International Conference on Data Mining*.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press, 1 edition.
- Wu, T., Chen, Y., and Han, J. (2007). Association mining in large databases: A re-examination of its measures. In *PKDD*, pages 621–628.
- Wu, T., Chen, Y., and Han, J. (2010). Re-examination of interestingness measures in pattern mining: a unified framework. *Data Min. Knowl. Discov.*, 21(3):371–397.