

Dinâmica das Relações de Coautoria nos Programas de Pós-Graduação em Computação no Brasil

Luciano A. Digiampietri¹, Jesús P. Mena-Chalco², Gabriela S. Silva¹,
Leonardo B. Oliveira³, Ana Paula Malheiros⁴, Dania Meira⁵

¹Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

²Centro de Matemática, Computação e Cognição da Universidade Federal do ABC

³Instituto de Ciências Exatas da Universidade Federal de Minas Gerais

⁴Elabora Consultoria e Participações

⁵Instituto de Matemática, Estatística e Computação Científica – Unicamp

digiampietri@usp.br

Abstract. *This paper describes the basis for a study of the dynamics relations of (co-)authorship among full time professors from the Brazilian Computer Science graduate programs. 889 researchers were identified, working in 45 graduate programs. A robust entity resolution heuristics were developed, allowing the identification of (co-)authorship relations among researchers with accuracy above 96%. The analysis of the social network allows finding interesting phenomena about the dynamics of the Brazilian scientific production, related to the increasing in the production inside and among the graduate programs.*

Resumo. *Este trabalho descreve as bases para um estudo da dinâmica de relações de coautoria entre pesquisadores associados aos programas de pós-graduação em Ciência da Computação avaliados pela CAPES no triênio 2007-2009. Ao todo, foram identificados 889 pesquisadores permanentes nos 45 programas de pós-graduação avaliados. Uma heurística robusta de resolução de entidades foi desenvolvida, possibilitando a identificação das relações de coautoria entre pesquisadores, com uma taxa de acerto superior a 96%. Através da observação das redes de coautoria é possível observar fenômenos interessantes da dinâmica da pesquisa brasileira, especialmente relacionados ao aumento da produção conjunta inter e intra programas de pós-graduação.*

1. Introdução

Nos últimos anos, a ciência Brasileira tem evoluído nos indicadores internacionais de produção científica¹. O trabalho conjunto dentro dos programas de pós-graduação e, tão ou mais importante que isso, entre programas de pós-graduação é um fator importante para se garantir uma pesquisa de qualidade, com nível internacional e, além disso, otimizando recursos.

¹Em 2011, no ranking acadêmico de universidades mundiais (ARWU), sete universidades brasileiras foram consideradas entre as 500 instituições de ensino superior mais valorizadas: <http://www.shanghairanking.com/ARWU2011.html>

Segundo indicadores obtidos da análise de um conjunto considerável de mais de 1 milhão de currículos Lattes, nas últimas décadas, a produção brasileira em periódicos internacionais está crescendo exponencialmente (entre 10% e 13% ao ano [Digiampietri et al. 2012]), bem como a interação acadêmica entre os pesquisadores. Um caso particular de interação entre pesquisadores refere-se à rede de colaboração acadêmica na forma de coautoria. O estudo de redes de coautoria e sua dinâmica (evolução) ao longo do tempo são tópicos que nos recentes anos estão recebendo grande interesse pela comunidade de acadêmica, pois permitem investigar e descobrir conhecimento relacionado com o comportamento social entre pesquisadores/grupos acadêmicos.

Este artigo visa apresentar a evolução das redes de coautoria, focando apenas nos docentes permanentes de programas de pós-graduação brasileiros na área de Ciência da Computação. Além de apresentar as redes formadas por estes docentes, também são analisadas, ao longo dos anos, as redes de coautoria entre esses programas de pós-graduação. Nesse sentido, são analisadas as produções bibliográficas dos docentes pertencentes ao corpo permanente dos 45 programas de pós-graduação em Ciência da Computação que possuem doutorado ou mestrado acadêmico, avaliados pela CAPES no triênio 2007-2009. Enquanto a seleção dos docentes utilizou este período de 2007 a 2009, as informações que são analisadas neste artigo envolvem uma janela de 30 anos, de 1981 a 2010. Acreditamos que, tenham sido consideradas produções bibliográficas de períodos além do triênio 2007-2009, as informações aqui obtidas refletem de forma aproximada o modo de interação acadêmica entre pesquisadores/programas associados à Ciência da Computação.

Neste artigo foram consideradas apenas as produções bibliográficas cadastradas por pesquisadores na Plataforma Lattes. Os currículos da Plataforma Lattes formam uma vasta fonte de informação individual sobre produção científica e tecnológica, porém, para identificação das relações (por exemplo, de coautoria, participação conjunta em projetos de pesquisa ou de relações de orientação acadêmica) entre diferentes pesquisadores se faz necessário o desenvolvimento de algoritmos para a resolução de entidades. Conforme será apresentado na Seção 2, estes algoritmos podem tirar vantagem da estruturação dos dados nos currículos Lattes para obter altas taxas de acertos na identificação de entidades.

O restante deste artigo está organizado da seguinte maneira. A Seção 2 apresenta a metodologia utilizada neste trabalho. A Seção 3 contém os resultados produzidos. A Seção 4 apresenta os principais trabalhos correlatos. Por fim, a Seção 5 contém as conclusões e os trabalhos futuros.

2. Metodologia

A metodologia adotada baseia-se na análise de produções bibliográficas de pessoas cadastradas na Plataforma Lattes e está composta de quatro processos: (i) identificação e obtenção de currículos Lattes, (ii) organização da informação, (iii) identificação das relações de coautoria, e (iv) produção das redes de coautoria.

2.1. Identificação e Obtenção dos Currículos

Para obter os currículos acadêmicos da Plataforma Lattes foi usado o *scriptLattes* [Mena-Chalco and Cesar-Jr. 2009] que utiliza como entrada uma lista com o identificador numérico dos currículos. Para a obtenção desta lista, foi feita uma consulta no

portal da CAPES² para a identificação de todos os programas de pós-graduação na área de computação que foram avaliados pela CAPES no triênio 2007-2009 e que possuem doutorado ou mestrado acadêmico. Esta consulta identificou 45 programas. Manualmente, foram examinados os *Cadernos de Indicadores*, que os programas de pós-graduação enviaram para a avaliação trienal, a fim de extrair a lista de docentes do corpo permanente destes programas. Ao todo, foram identificados 889 professores permanentes.

A lista com os identificadores numéricos de currículos foi passada como parâmetro de entrada para o *scriptLattes* que baixou os arquivos HTML de cada currículo.

2.2. Organização da Informação

Dada a utilização dos currículos Lattes em formato HTML, onde as partes constitutivas de uma produção bibliográfica não são discriminadas, foi necessário o desenvolvimento de um *parser* para extrair os artigos completos publicados em periódicos. Para cada artigo, os seguintes campos foram extraídos: título, páginas, volume, autores, ano e local.

Adicionalmente, cada artigo identificado foi armazenado em um banco de dados relacional juntamente com um identificador único, identificador numérico Lattes do pesquisador, nome do programa de pós-graduação e nome do pesquisador possuidor do currículo. Esta forma de armazenamento permite uma rápida e flexível consulta de produções bibliográficas: (i) publicadas em diferentes períodos de tempo; e (ii) agrupadas por nome de programas de pós-graduação.

2.3. Identificação das Relações

Neste artigo foi utilizada como relação entre pesquisadores a coautoria de artigos publicados em periódicos científicos. Para a identificação das relações, foi desenvolvida uma heurística de resolução de entidades. Dois artigos são considerados o mesmo (a mesma entidade) se três condições forem satisfeitas: (i) os títulos forem compatíveis; (ii) a lista de autores for compatível; e (iii) as demais informações forem compatíveis. A seguir cada uma dessas condições é descrita:

- **Condição 1:** Dois títulos de produções bibliográficas são considerados compatíveis se são iguais; *OU* se a diferença entre o tamanho dos dois títulos for menor do que um terço da soma do tamanho dos títulos E {ambos possuem mais de 10 caracteres e um estiver contido dentro do outro *OU* a Distância Levenshtein [Levenshtein 1966] entre os dois títulos for menos do que 5}. Obviamente a última parte da condição garantiria a primeira parte, porém, devido à maior complexidade computacional necessária para calculá-la, a verificação de compatibilidade de título é executada na ordem apresentada.
- **Condição 2:** Duas lista de (co)autores são consideradas compatíveis se, ao se comparar as duas listas, houver mais autores em comum do que diferentes, considerando-se apenas o casamento exato do último sobrenome de cada autor. Foi considerado o último sobrenome por, na maioria dos casos, ser invariante a abreviações.
- **Condição 3:** As demais informações serão compatíveis se ao menos dois dos seguintes quatro campos forem iguais: ano de publicação, local, páginas e volume.

²<http://www.capes.gov.br/cursos-recomendados>

Este algoritmo foi desenvolvido, testado e calibrado sobre um conjunto pequeno de dados, formado por 5 pesquisadores e 330 itens de produção bibliográfica. Este conjunto foi escolhido por conter diversas publicações em coautorias de forma a facilitar a identificação das características que indicam se diferentes registros de publicações correspondem à mesma produção. Conforme visto na descrição da heurística, este utilizou apenas os campos comuns às produções bibliográficas presentes nos Currículos Lattes. A Tabela 1 discrimina os itens do conjunto de dados de acordo com o tipo de produção.

Tabela 1. Conjunto de Dados Utilizado no Treinamento

Tipo de Produção Bibliográfica	Quantidade
Artigos completos publicados em periódicos	90
Artigos aceitos para publicação	4
Trabalhos completos publicados em anais de congressos	179
Resumos expandidos publicados em anais de congressos	22
Resumos publicados em anais de congressos	17
Livros publicados organizados ou edições	5
Capítulos de livros publicados	13
Total	330

Com o algoritmo calibrado um banco de dados de publicações foi verificado manualmente para que a precisão do algoritmo pudesse ser verificada. Neste banco de dados foram inseridos todos os artigos completos publicados em periódicos dos docentes permanentes no triênio 2007-2009 do programa de pós-graduação em Ciências da Computação do Instituto de Matemática e Estatística da Universidade de São Paulo³. Foram analisados os currículos de 36 docentes, totalizando 620 registros de publicações. A verificação manual permitiu observar que destes 620 registros são referentes a 486 artigos distintos, sendo que 374 não possuíam dois coautores dentro do conjunto de docentes analisados.

O algoritmo desenvolvido foi testado e validado utilizando este conjunto dados, lembrando que o objetivo do algoritmo é resolver entidades (verificar se dois registros diferentes se referem a mesma entidade, no caso, ao mesmo artigo publicado em revista científica). Os resultados desta validação foram os seguintes: dos 486 artigos diferentes existentes, o algoritmo identificou corretamente 468 (taxa de verdadeiros positivos igual a 96,3% do total de artigos). O algoritmo também identificou 5 artigos como únicos quando na verdade eram artigos diferentes (falsos positivos) e deixou de unir 36 registros que correspondiam a 18 artigos diferentes (falsos negativos).

Com base nestes resultados, o algoritmo foi considerado robusto o suficiente para ser utilizado na identificação das relações de coautoria para a formação das redes sociais.

2.4. Produção das Redes Sociais

Dos 889 pesquisadores analisados, 743 haviam publicado um ou mais artigos completos em periódicos no período de 1981 a 2010. Um total de 8.611 artigos em periódicos foram identificados, sendo que 2.411 são resultantes de coautorias entre membros de um mesmo programa de pós-graduação e 890 são resultantes de coautorias entre membros de diferentes programas de pós-graduação em Ciência da Computação.

³<http://www.ime.usp.br/>

Com base nas relações de coautorias dois conjuntos de redes foram produzidos. Em um deles, cada docente é um elemento da rede e as relações entre docentes são dados por suas coautorias. No outro conjunto, cada programa de pós-graduação é um elemento da rede e a coautoria entre programas (extraída da coautoria entre membros dos programas) é utilizada como relação. A Tabela 2 descreve todas as redes geradas. Para a visualização de cada rede, o tamanho de cada elemento da rede é proporcional ao seu *Author Rank*⁴ [Liu et al. 2005]. Na próxima seção serão apresentadas e analisadas as principais redes produzidas.

Tabela 2. Discriminação das 156 Redes de Coautorias Produzidas

	Entre Docentes	Entre Programas
Ano a Ano	30	30
Anual Acumulativa	30	30
Triênio a Triênio	9	9
Trienal Acumulativa	9	9

3. Resultados

Esta seção apresenta algumas das redes de coautorias produzidas, uma breve análise sobre elas e algumas estatísticas. Pelo fato das redes terem sido produzidas em uma janela de 30 anos e isso gerar um volume muito grande de gráficos, neste artigo serão apresentados apenas dados trienais. As dinâmicas das redes de docentes e de programas, ano a ano, podem ser vistas através de dois Gifs animados disponível em: www.each.usp.br/digiampietri/redes/docentes.gif e www.each.usp.br/digiampietri/redes/programas.gif.

As Figuras 1 e 2 apresentam redes de docentes, onde cada nó representa um docente permanente dos programas de pós-graduação em computação avaliados pela CAPES no triênio 2007-2009. Os nós estão coloridos de acordo com o programa que o docente pertence e o tamanho do nó é proporcional ao *Author Rank* do docente, considerando apenas as publicações em periódico analisadas. As relações entre os nós são definidas por relações de coautoria.

A Figura 1 contém as coautorias entre docentes nos triênios de 1992-1994 até 2007-2009. É possível observar o crescimento, triênio a triênio na colaboração entre docentes, bem como a manutenção/regularidade de valores altos de *author rank* de alguns docentes ao longo de todo o período observado. Diversas pequenas cliques podem ser observadas nestas redes, formadas principalmente por integrantes de um mesmo programa. Além disso, é possível observar um grande aumento nas coautorias ao longo dos triênios e aumento no número de componentes fortemente conexos e de cliques nos grafos correspondentes a estas redes sociais.

A Figura 2 contém as coautorias acumuladas entre docentes. A análise utilizou as publicações a partir de 1981 e as redes com o total acumulado dessas publicações e apresentadas nestas figuras variam de 1981-1994 até 1981-2009. Com esta visão acumulada das coautorias é possível verificar a forte conexão entre docentes de um mesmo programa e docentes de diferentes programas. Em especial, ao se observar as redes da

⁴A medida de *Author Rank* é, comumente, utilizada para investigar o grau de colaboração de um pesquisador com os outros pesquisadores do mesmo grupo. Quanto maior a colaboração, maior o valor de *Author Rank*.

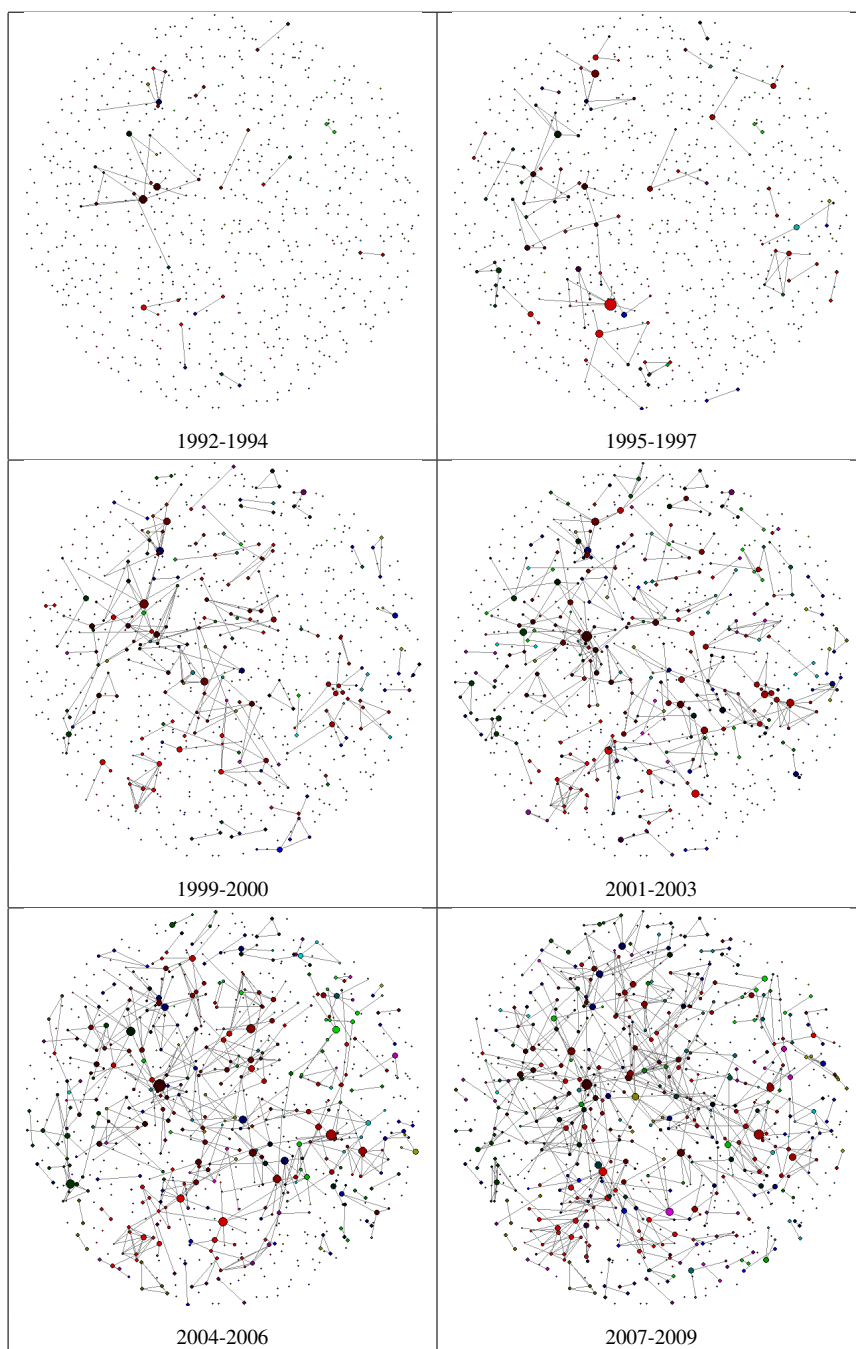


Figura 1. Coautorias entre Docentes

Figura 1, triênio 2007-2009, é possível notar que praticamente todos os nós de maior tamanho possuem uma forte ligação com ambos: docentes do mesmo programa e docentes de outro programa, sendo, provavelmente, ligações entre grupos de pesquisa de diferentes programas.

As Figuras 3 e 4 apresentam redes de programas de pós-graduação, onde cada nó representa um programa em computação avaliado pela CAPES no triênio 2007-2009. O tamanho de cada nó é proporcional ao *Author Rank* do programa, considerando apenas as publicações em periódico analisadas. As relações entre os nós são definidas por relações

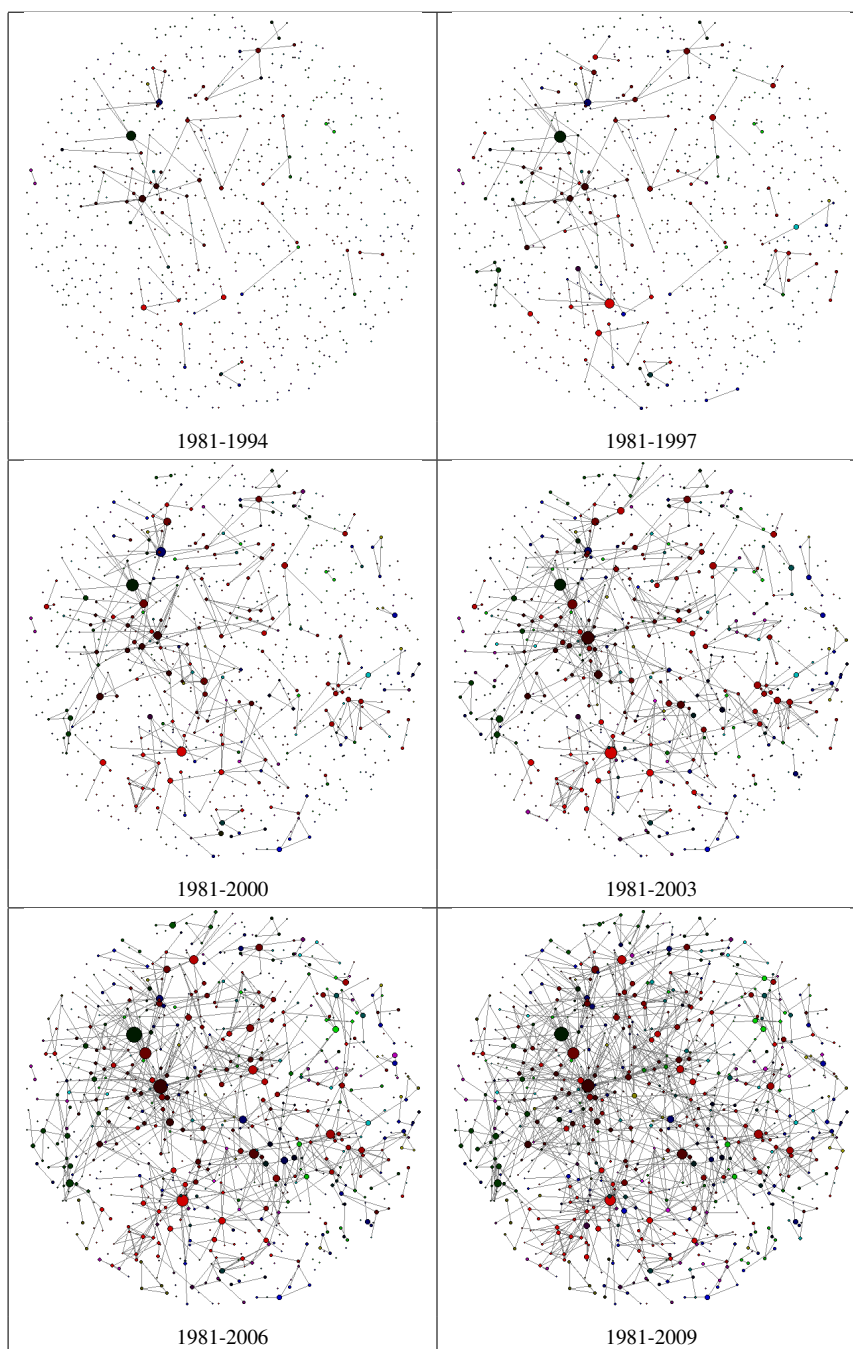


Figura 2. Coautorias Acumuladas entre Docentes

de coautoria (entre os docentes dos diferentes programas).

A Figura 3 contém as coautorias entre programas dos triênios de 1992-1994 a 2007-2009. As duas características mais marcantes deste conjunto de redes são: a grande evolução no número de colaborações ao longo dos triênios e o fato de alguns programas serem o ponto de conexão com diversos outros, conforme pode ser observado nos programas que apresentam um maior tamanho nas redes apresentadas.

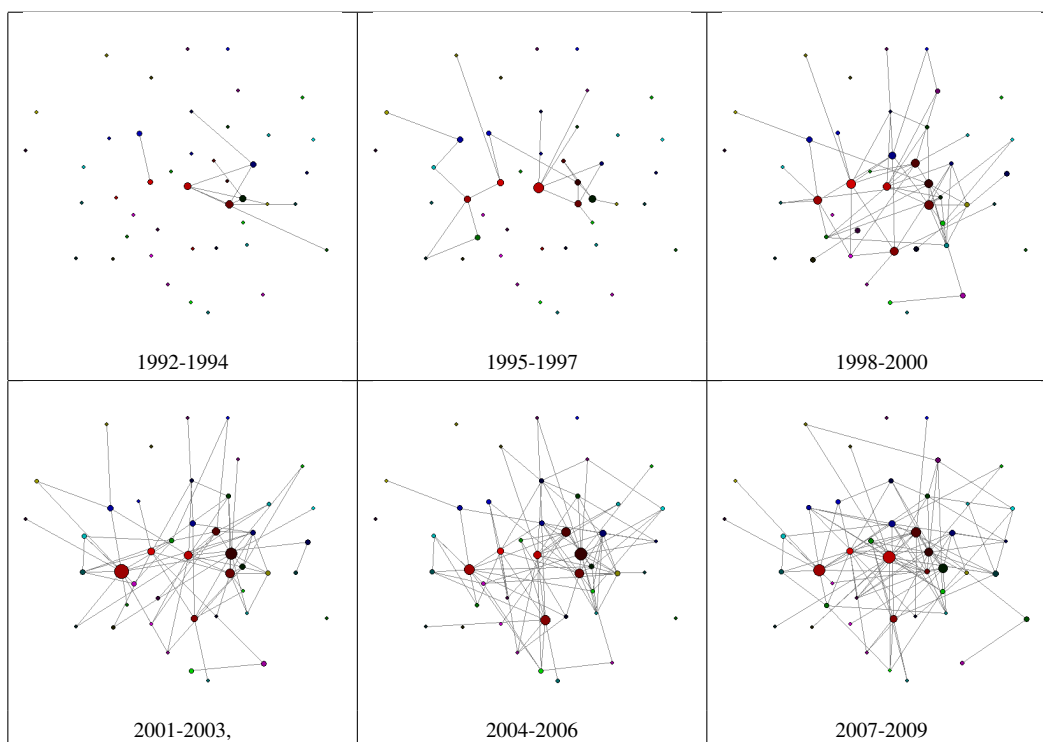


Figura 3. Coautorias entre Programas

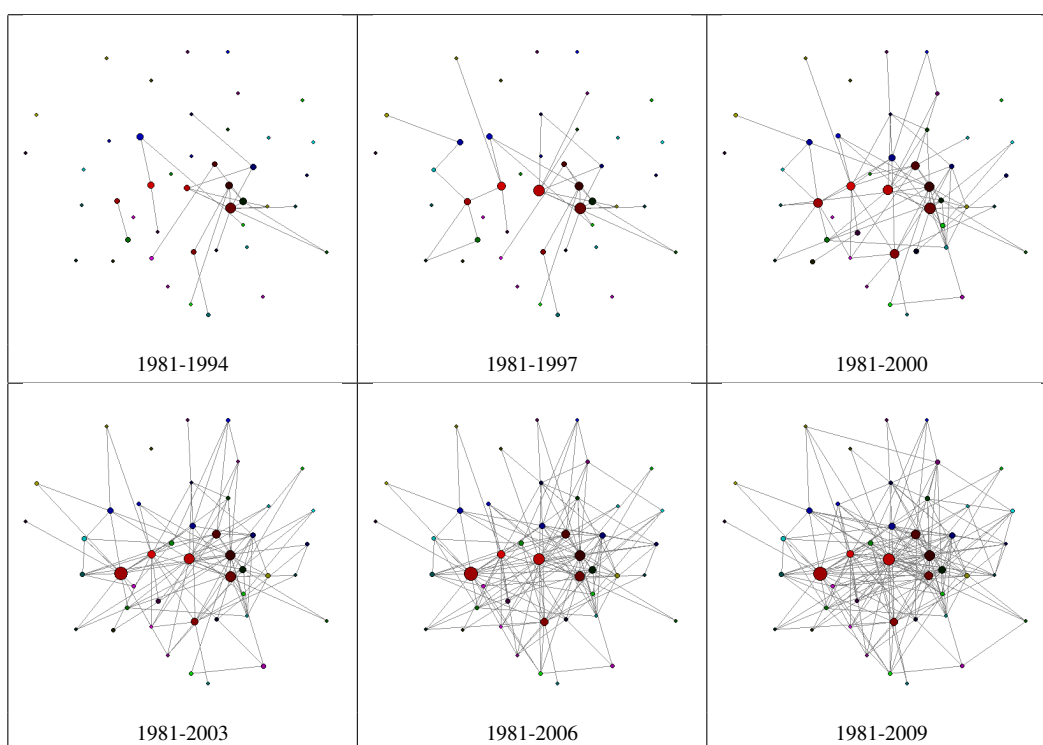


Figura 4. Coautorias Acumuladas entre Programas

A Figura 4 contém as coautorias acumuladas entre programas nos períodos de 1981-1994 a 1981-2009. É possível verificar que as últimas redes forma um único com-

ponente conexo. Ao se analisar os números relacionados às publicações que foram utilizadas como base para a formação da rede, algumas características que chamam a atenção nas redes produzidas podem ser facilmente confirmadas.

A Figura 5 apresenta um gráfico contendo o número de artigos publicados em periódicos em cada um dos triênios analisados e a relação entre artigos em coautoria entre membros dos programas e o número total de publicações desses membros. É possível verificar que em quase todo período analisado o número de publicações cresceu de maneira exponencial. Além disso, no período de 1992 a 2006 a relação do número de publicações em coautoria pelo número total de publicações cresceu (considerando tanto as coautorias entre docentes quanto entre programas), indicando uma maior colaboração entre docentes e programas. Esta característica não pode ser observada no último triênio avaliado, mas um dos possíveis motivos é a falta de atualização dos currículos: para que um artigo entre na lista de artigos publicados no triênio basta que um dos seus autores tenha inserido este artigo no currículo, porém, para ser considerada uma coautoria dois ou mais autores precisam ter inserido essa informação em seus currículos. De qualquer maneira, este evento precisa ser investigado de maneira mais profunda.

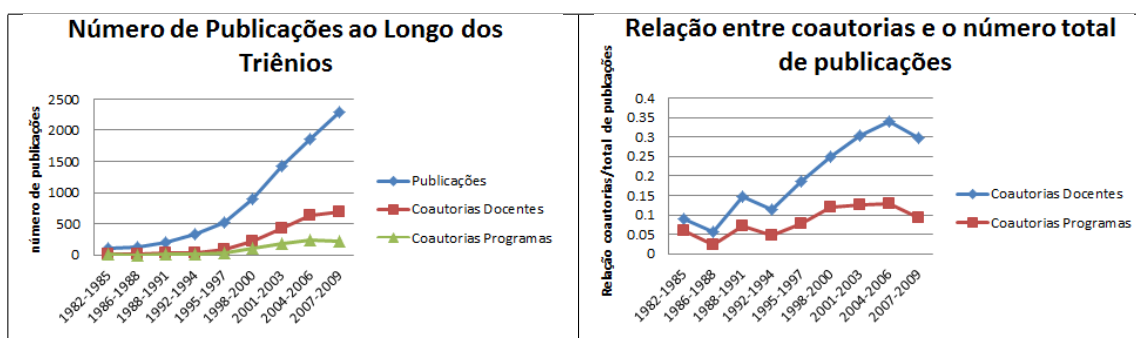


Figura 5. Número de Publicações e Relação entre as Coautorias e o Total de Publicações por Triênio

4. Trabalhos Correlatos

Dois conjuntos distintos de trabalhos correlatos foram analisados, um com enfoque na evolução ou dinâmica das redes sociais e outro de trabalhos que utilizaram currículos da Plataforma Lattes. A maioria dos trabalhos correlatos que trabalham com a evolução ou dinâmica de redes sociais foca em redes específicas.

Por exemplo, Horn *et al.* [Horn et al. 2004] avaliaram a evolução e o impacto das redes de coautoria relacionadas ao assunto “trabalho cooperativo apoiado pelo computador”, identificando quais as principais áreas correlatas e procurando por padrões de colaboração. Outro exemplo deste tipo de enfoque é o trabalho de Hayat e Lyons [Hayat and Lyons 2010] que analisaram as redes de coautoria formadas pelos autores que publicam na conferência CASCON. Através do uso da análise de redes sociais, os autores identificaram características e padrões na rede de forma a sugerirem direções para facilitar o desenvolvimento de comunidades ou eventos científicos. Sharma e Urs [Sharma and Urs 2008] analisaram a rede de coautorias de pesquisadores que publicam no tema “bibliotecas digitais”. Partindo dos editores e dos autores mais conhecidos, elas montaram e analisaram as redes de coautoria de forma a detalhar esta rede específica.

Guo *et al.* [Guo et al. 2009] analisam a produção e divulgação de conteúdos em diferentes tipos de redes sociais online. Os autores conseguiram identificar diferentes padrões temporais na produção, qualidade e esforço dos conteúdos.

Outros trabalhos estão focados em apresentar modelos e infraestruturas mais gerais para a visualização e análise da dinâmica das redes sociais. Berger Wolf e Saia [Berger-Wolf and Saia 2006] apresentaram um modelo formal para permitir a análise dinâmica das redes. Baumes *et al.* [Baumes et al. 2008] apresentam o ambiente *ViSAGE*, que é um sistema para simulação e modelagem de redes sociais dinâmicas e visa a facilitar a teorização e validação de propriedades de uma dada rede. Bin Wu *et al.* [Wu et al. 2009] desenvolveram *Group CMR* uma infraestrutura para a análise das ligações de clientes aos *call centers*. Nesta infraestrutura os clientes são agrupados em grupos de acordo com alguns padrões identificados, o que viabiliza a análise de quantidades massivas de dados. Kang *et al.* [Kang et al. 2007] desenvolveram uma ferramenta chamada *C-Group* para facilitar a visualização dinâmica de redes. O principal diferencial da ferramenta desenvolvida é permitir o acompanhamento de pares de elementos (e não apenas de indivíduos, como ocorre na maioria dos sistemas similares).

Mena-Chalco e Cesar Junior [Mena-Chalco and Cesar-Jr. 2009] desenvolveram o *scriptLattes*, um sistema de código livre para a extração e organização de dados de currículos Lattes. A partir de uma lista de identificadores de currículos, o sistema baixa e organiza esses currículos de forma a permitir uma visão organizada das informações do conjunto de docentes, bem como exibe o grafo de coautorias. Uma análise sobre interação considerando coautorias realizadas entre 2000 e 2010 dos pesquisadores de 4 áreas de conhecimento usando a mesma ferramenta foi discutida em [Mena-Chalco and Cesar-Jr. 2011]. Alves *et al.* [Alves et al. 2011] desenvolveram o sistema *LattesMinner* para a extração de dados da Plataforma Lattes. Este sistema não extrai todas as informações disponíveis nos currículos, mas extrai e organiza uma quantidade relevante de informação (por exemplo, informações sobre artigos publicados em periódicos e artigos completos publicados em conferências). O sistema foi desenvolvido para servir de entrada para sistemas de análise ou visualização de redes sociais. Laender *et al.* [Laender et al. 2011] descrevem parte do projeto *CiênciaBrasil* que visa a fornecer instrumentos para facilitar a visualização e o entendimento da produção científica brasileira, sendo que uma das ferramentas é responsável por exibir a rede de coautorias de um dado autor, com base nas informações de seu currículo Lattes.

Assim como diversos trabalhos correlatos, neste artigo foi analisado um grupo específico de pesquisadores (aqueles que são docentes permanentes em programas de pós-graduação em computação no Brasil). Além da análise das redes dos docentes em um período de 30 anos, neste artigo também foram analisadas as redes formadas pelos programas de pós-graduação. Enquanto a análise de um novo grupo por si só já possui relevância, neste artigo foi desenvolvido um novo algoritmo de resolução de entidades específico para tratar dados de publicações de currículos da Plataforma Lattes, o que permite que as análises realizadas sejam feitas de maneira mais precisa considerando dois eixos importantes: pesquisadores e programas de pós-graduação.

5. Conclusões e Trabalhos Futuros

Este artigo apresentou um conjunto de redes de coautoria formadas por dados extraídos de currículos da Plataforma Lattes. Os dados utilizados foram os currículos dos docentes permanentes de programas de pós-graduação em Ciência da Computação no Brasil, avaliados pela CAPES no triênio 2007-2009 e que possuíam doutorado ou mestrado acadêmico. As relações utilizadas para a formação das redes sociais foram as relações de coautoria de artigos completos publicados em periódicos.

Dois conjuntos de redes foram analisados. No primeiro, as redes eram formadas por docentes e as relações consideradas foram as coautorias entre estes docentes. No segundo, as redes eram formadas por programas de pós-graduação em computação e as relações consideradas foram as coautorias entre os docentes desses programas.

Para o estabelecimento das relações, foi desenvolvido um algoritmo para resolução de entidades para identificar diferentes itens dos currículos Lattes correspondentes à mesma publicação. O algoritmo desenvolvido foi testado sobre um conjunto de dados anotado manualmente e obteve uma taxa de acerto superior a 96%.

Pelas análises das redes produzidas foi possível observar um fortalecimento gradual das coautorias tanto entre docentes quanto entre programas. Foi possível também identificar que poucos docentes exercem a função de ligação entre diferentes conjuntos de pesquisadores de diversos programas. Esta característica pode indicar que uma boa política para o fortalecimento dos grupos de pesquisa nacionais, bem como para o intercâmbio de conhecimentos é incentivar projetos de professores visitantes entre programas nacionais (e não apenas visitantes de fora do país).

Como trabalhos futuros pretende-se analisar a influência das relações de orientação nas redes de coautoria, detalhe a função que os principais docentes de cada programa exercem para o desenvolvimento da rede formada pelo programa, bem como das relações entre diferentes programas. Além disso, pretende-se estudar as relações temporais de coautoria e vinculá-las aos projetos de pesquisa presentes nos dados dos currículos Lattes de forma a tentar prever mudanças na dinâmica das redes sociais de pesquisadores.

Agradecimentos

O trabalho apresentado neste artigo foi parcialmente financiado pelo Programa de Educação Tutorial (MEC/SESu).

Referências

- Alves, A. D., Yanasse, H. H., and Soma, N. Y. (2011). LattesMiner: a multilingual dsl for information extraction from lattes platform. In *Proceedings of SPLASH'11, SPLASH '11 Workshops*, pages 85–92, New York, NY, USA. ACM.
- Baumes, J., Chen, H.-C. J., Francisco, M., Goldberg, M., Magdon-Ismail, M., and Wallace, W. (2008). Visage: A virtual laboratory for simulation and analysis of social group evolution. *ACM Trans. Auton. Adapt. Syst.*, 3(3):8:1–8:35.
- Berger-Wolf, T. Y. and Saia, J. (2006). A framework for analysis of dynamic social networks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 523–528, New York, NY, USA.

- Digiampietri, L., Mena-Chalco, J., de Jesús Pérez-Alcázar, J., Tuesta, E. F., Delgado, K., and Mugnaini, R. (2012). Minerando e caracterizando dados de currículos lattes. In *CSBC 2012 - BraSNAM*.
- Guo, L., Tan, E., Chen, S., Zhang, X., and Zhao, Y. E. (2009). Analyzing patterns of user content generation in online social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '09*, pages 369–378, New York, NY, USA. ACM.
- Hayat, Z. and Lyons, K. (2010). The evolution of the cascon community: a social network analysis. In *Proceedings of the 2010 Conference of the Center for Advanced Studies on Collaborative Research, CASCON '10*, pages 1–12, Riverton, NJ, USA. IBM Corp.
- Horn, D. B., Finholt, T. A., Birnholtz, J. P., Motwani, D., and Jayaraman, S. (2004). Six degrees of jonathan grudin: a social network analysis of the evolution and impact of cscw research. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work, CSCW '04*, pages 582–591, New York, NY, USA. ACM.
- Kang, H., Getoor, L., and Singh, L. (2007). Visual analysis of dynamic group membership in temporal social networks. *SIGKDD Explor. Newsl.*, 9(2):13–21.
- Laender, A. H., Moro, M. M., Gonçalves, M. A., Davis, Jr., C. A., da Silva, A. S., Silva, A. J., Bigonha, C. A., Dalip, D. H., Barbosa, E. M., Cortez, E., Procópio, Jr., P. S., de Alencar, R. O., Cardoso, T. N., and Salles, T. (2011). Building a research social network from an individual perspective. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries, JCDL '11*, pages 427–428, New York, NY, USA. ACM.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8):707–710.
- Liu, X., Bollen, J., Nelson, M., and de Sompel, H. V. (2005). Co-authorship networks in the digital library research community. *Informations Processing and Management*, 41(6):1462–1480.
- 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.
- Sharma, M. and Urs, S. R. (2008). Network dynamics of scholarship: a social network analysis of digital library community. In *Proceedings of the 2nd PhD workshop on Information and knowledge management*, pages 101–104, New York, NY, USA.
- Wu, B., Ye, Q., Yang, S., and Wang, B. (2009). Group crm: a new telecom crm framework from social network perspective. In *Proceedings of the 1st ACM international workshop on Complex networks meet information & knowledge management, CNIKM '09*, pages 3–10, New York, NY, USA. ACM.