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

BrASNAM

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

ANALYSIS OF THE EVOLUTION OF COAUTHORSHIP IN COMPUTING GRADUATE PROGRAMS IN BRAZIL

(artigo submetido em março de 2013)

Luciano A. Digiampietri
Escola de Artes, Ciências e Humanidades - Universidade de São Paulo (USP)
digiampietri@usp.br

Jesús P. Mena-Chalco
Centro de Matemática, Computação e Cognição - Universidade Federal do ABC (UFABC)
jesus.mena@ufabc.edu.br

Gabriela S. Silva
Escola de Artes, Ciências e Humanidades - Universidade de São Paulo (USP)
gabriela.scardine.silva@usp.br

Leonardo B. Oliveira
Instituto de Ciências Exatas - Universidade Federal de Minas Gerais (UFMG)
leonardo.barbosa@dcc.ufmg.br

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

Ana Paula Malheiro
Elabora Consultoria e Participações
ana.paula@elabsis.com

Dania Meira
Instituto de Matemática, Estatística e Computação Científica
Universidade de Campinas (Unicamp)
meira.dania@gmail.com

ABSTRACT

This paper describes the basis for a study of the dynamic relationships of (co-)authorship among full time professors from Brazilian Computer Science graduate programs. 889 researchers were identified, working in 45 graduate programs. A robust entity resolution heuristic was developed, allowing the identification of (co-)authorship relationships among researchers with accuracy above 96%. The social network analysis allowed for the discovery of some interesting phenomena about the dynamics of the Brazilian scientific production, related to the increasing in the production within and among graduate programs.

Keywords: Lattes platform; social network; co-authorship network; data mining.

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%. Com base na análise das redes de coautoria foi possível observar fenômenos interessantes da dinâmica da pesquisa brasileira, relacionados especialmente ao aumento da produção conjunta inter e intra programas de pós-graduação.

Palavras-chave: plataforma Lattes; rede social; rede de coautoria; mineração de dados.

1 INTRODUÇÃO

Atualmente, é possível encontrar na Web uma grande quantidade de dados referentes aos mais diversos assuntos. Dentre esses dados estão informações relevantes sobre os pesquisadores, tais como suas publicações científicas, informações sobre projetos de pesquisa de que participam e até mesmo seus currículos.

Nos últimos anos, a ciência brasileira tem evoluído nos indicadores internacionais de produção científica[1]. O trabalho colaborativo dentro dos programas de pós-graduação e entre programas de pós-graduação é um fator importante para garantir uma pesquisa de qualidade, com nível internacional e, além disso, a otimização de recursos.

Segundo indicadores obtidos da análise de um conjunto de mais de um milhão de currículos Lattes, nas últimas décadas, a produção brasileira em periódicos internacionais está crescendo de forma consistente, entre 10% e 13% ao ano, conforme Digiampietri *et al.* (2012a). Também cresce 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) recentemente está recebendo grande interesse da comunidade acadêmica, pois permite investigar e adquirir conhecimento relacionado com o comportamento social entre pesquisadores/grupos acadêmicos.

Este artigo visa a apresentar a evolução das redes de coautoria, focando 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 examinadas, ao longo dos anos, as redes de coautoria entre os 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 1983 a 2012. Acreditamos que, mesmo considerando produções bibliográficas além do triênio de 2007 a 2009, as informações aqui obtidas refletem de forma satisfatória 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

[1] Em 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).

(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 acerto na identificação de entidades.

O presente artigo revisita e estende o trabalho intitulado "Dinâmica das relações de coautoria nos programas de pós-graduação em computação no Brasil" (DIGIAMPIETRI *et al.*, 2012b) de maneira a realizar uma análise mais completa das relações de coautoria, considerando além dos artigos publicados em periódicos, os artigos completos publicados em conferências, forma de produção extremamente relevante na área de Ciência da Computação (MENA-CHALCO, DIGIAMPIETRI e OLIVEIRA, 2012). Além disso, a janela de avaliação foi atualizada, incluindo informações de 1983 a 2012. Por fim, novas métricas foram extraídas de forma a destacar a evolução das redes de coautoria.

O restante deste artigo está organizado da seguinte maneira. Na Seção 2, os principais trabalhos correlatos são apresentados. Na Seção 3 apresenta-se a metodologia considerada neste trabalho. Na Seção 4 são apresentados os resultados obtidos. Por fim, na Seção 5 são apresentadas as conclusões e indicados possíveis trabalhos futuros.

2 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.* (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 (2010) que analisaram as redes de coautoria formadas pelos autores que publicam na conferência CASCON. Por meio de 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 (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*. (2009) analisam a produção e divulgação de conteúdos em diferentes tipos de redes sociais *online*, conseguindo 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 (2006) apresentam um modelo formal para permitir a análise dinâmica das redes.

Baumes *et al*. (2008), por sua vez, apresentam o ambiente *ViSAGE*, que é um sistema para simulação e modelagem de redes sociais dinâmicas que visa a facilitar a teorização e validação de propriedades de uma dada rede.

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 de acordo com alguns padrões identificados, o que viabiliza a análise de quantidades massivas de dados.

Kang *et al*. (2007) desenvolveram uma ferramenta chamada *C-Group* para facilitar a visualização dinâmica de redes. O principal diferencial da ferramenta 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 (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 quatro áreas de conhecimento usando a mesma ferramenta foi discutida em Mena-Chalco e Cesar-Jr, 2011).

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.* (2011) descrevem parte do projeto *CiênciaBrasil* que visa a fornecer instrumentos para facilitar a visualização e o entendimento da produção científica brasileira. Uma das ferramentas é responsável por exibir a rede de coautorias de um dado autor, com base nas informações de seu currículo Lattes.

Mena-Chalco *et al*. (2014) utilizaram mais de um milhão de currículos da Plataforma Lattes para criar redes de coautorias visando a caracterizar a interação acadêmica no Brasil nas diferentes áreas de conhecimento.

Digiampietri *et al.* (2014) realizaram uma análise profunda de diferentes índices de produtividade dos docentes pertencentes aos programas de pós-graduação em Ciência da Computação no Brasil e conseguiram correlacionar métricas estruturais das redes de coautoria com as notas atribuídas pela CAPES aos programas de pós-graduação.

3 METODOLOGIA

A metodologia adotada baseia-se na análise de produções bibliográficas de professores/pesquisadores cadastrados na Plataforma Lattes e inclui 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.

3.1 IDENTIFICAÇÃO E OBTENÇÃO DOS CURRÍCULOS

Para obter os currículos acadêmicos da Plataforma Lattes foi usado o *scriptLattes* (MENA-CHALCO e CESAR-JR, 2009). O *scriptLattes* recebe como entrada uma lista com o identificador numérico dos currículos. Para a obtenção desta lista, foi feita uma consulta no portal da CAPES[2] 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. 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 o arquivo HTML de cada currículo.

3.2 ORGANIZAÇÃO DA INFORMAÇÃO

Como as partes constitutivas de uma produção bibliográfica não são discriminadas nos currículos Lattes em formato HTML, foi necessário o desenvolvimento de um *parser* para extrair tais partes (dos artigos completos publicados em conferências ou 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 o identificador numérico Lattes do pesquisador, o nome do programa de pós-graduação e o nome do pesquisador possuidor do currículo. Esta forma de armazenamento permite uma rápida e flexível consulta de produções bibliográficas: (i) publi-

[2] http://www.capes.gov.br/cursos-recomendados

cadas em diferentes períodos de tempo; e (ii) agrupadas pelos programas de pós-graduação.

3.3 IDENTIFICAÇÃO DAS RELAÇÕES

Neste artigo foi utilizada como relação entre pesquisadores a coautoria de artigos. 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 menor 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 foi executada na ordem apresentada.
- **Condição 2:** Duas listas 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.

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, foram utilizados 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

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**

Fonte: elaborada pelos autores a partir de dados da pesquisa

Com o algoritmo calibrado, um banco de dados de publicações foi anotado 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 Paulo[3]. Foram analisados os currículos de 36 docentes, totalizando 620 registros de publicações. A verificação manual permitiu observar que estes 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 de dados, lembrando que o objetivo do algoritmo era resolver entidades (verificar se dois registros diferentes se referem à mesma entidade, no caso, ao mesmo artigo científico). 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.

3.4 PRODUÇÃO DAS REDES SOCIAIS

Dos 889 pesquisadores analisados, 859 haviam publicado um ou mais artigos completos em conferências ou periódicos no período de 1983 a 2012 em coautoria com outro docente do conjunto analisado. Ao todo, 52.075 artigos completos foram produzidos pelos docentes no período,

[3] http://www.ime.usp.br/

sendo 10.669 publicados em periódicos e 41.406 em conferências. Deste total, 10.402 correspondem a coautorias entre os docentes analisados, sendo 5.370 referentes a coautorias entre membros de diferentes programas de pós-graduação em Ciência da Computação.

Com base nas relações de coautoria dois conjuntos de redes foram produzidos. Em um deles, cada docente é um elemento da rede e as relações entre docentes são dadas 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.

Tabela 2. Discriminação das 160 redes de coautorias produzidas

	Entre docentes	Entre programas
Ano a ano	30	30
Anual acumulativa	30	30
Triênio a triênio	10	10
Trienal acumulativa	10	10
Total	**80**	**80**

Fonte: elaborada pelos autores a partir de dados da pesquisa

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.

4 RESULTADOS

Esta seção apresenta algumas das redes de coautorias produzidas, uma breve análise sobre elas e algumas estatísticas. Pelo fato de as redes terem sido produzidas em uma janela de 30 anos e isso gerar um volume muito grande de gráficos, neste artigo são apresentados apenas alguns gráficos considerados mais informativos. As dinâmicas das redes de docentes e de programas, ano a ano, podem ser observadas por meio de dois *gifs* animados disponíveis em: http://each.uspnet.usp.br/digiampietri/redes/.

Antes da apresentação e discussão das redes é apresentado um resumo da evolução da produção dos docentes ao longo dos trinta anos avaliados. O aumento, em praticamente todos os anos destes índices é refletido na densidade de arestas das redes de coautoria.

Na Figura 1 apresenta-se a evolução da produção de artigos completos pelos docentes dos programas de pós-graduação em Ciência da Computação. Os artigos estão organizados em três categorias: coautorias apenas em um programa (quando apenas docentes de um mesmo pro-

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

grama fazem parte da lista de autores de um artigo); coautorias entre programas (quando há docentes de mais de um programa na lista de autores); e totais individuais (produção que, na lista de autores, possui apenas um docente dentre os 889 analisados). Nesta figura é possível observar que o número total de artigos tem aumentado praticamente todos os anos. Vale ressaltar que o termo "produção em coautoria" está sendo usado para verificar a coautoria endógena, i. e., colaboração apenas entre os 889 docentes analisados. Um artigo produzido por apenas um dos docentes deste grupo (mas que pode conter outros autores de fora deste grupo de 889 docentes) está sendo considerado como uma produção individual, nesta análise.

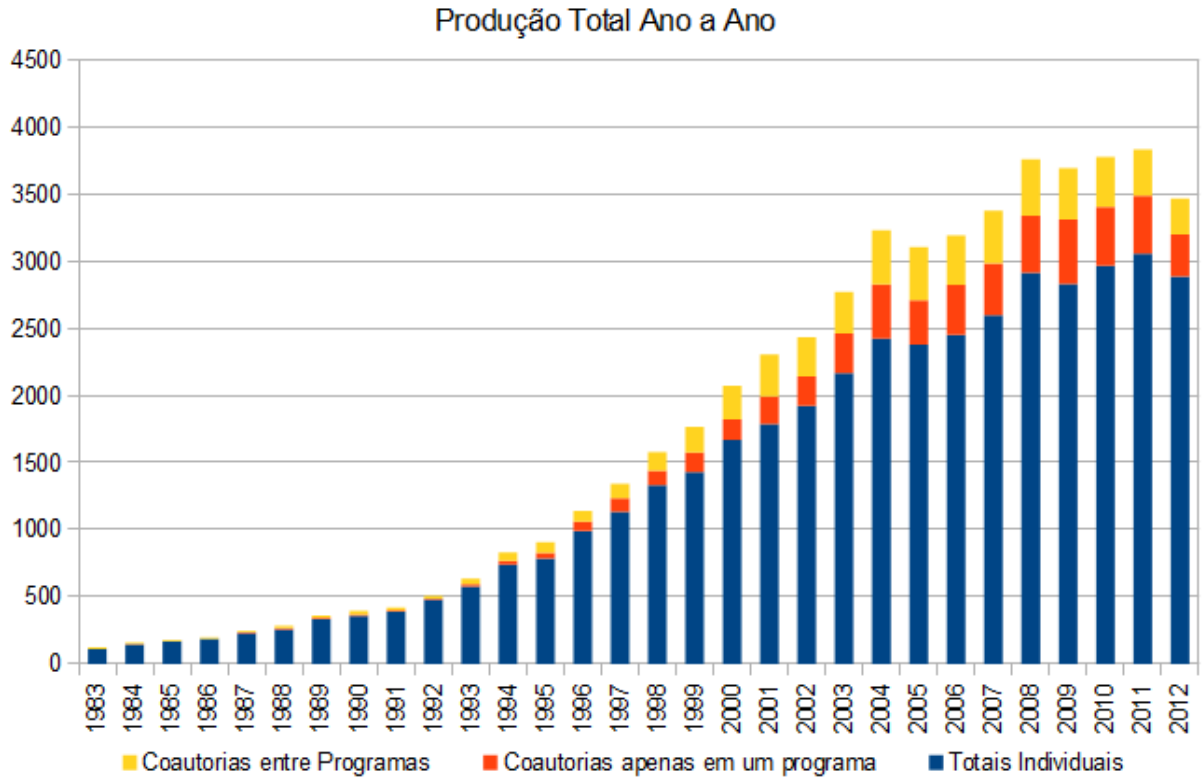

Figura 1. Evolução da produção ao longo dos anos

Fonte: elaborada pelos autores a partir de dados da pesquisa.

A Figura 2 apresenta a evolução das coautorias ao longo dos anos em relação ao número total de artigos publicados nesse período. É interessante observar que a relação entre as produções em coautoria e total de produções também tem aumentado nos últimos anos. Na média, no período de 1983 a 1992, apenas 7,65% dos artigos foram publicados em coautoria entre estes docentes. No período de 1993 a 2002 a média teve um incremento, atingindo 16,02% do total de artigos publicados. De 2003 a 2012 a porcentagem de produção em coautoria subiu para 22,08%. Nesta figura também é possível observar que, a partir do ano de 1996, as coautorias em periódicos tiveram um aumento relativo maior do que das coautorias em conferências, sendo que em 2004 e na maioria dos anos

subsequentes, proporcionalmente, foram identificadas mais coautorias nas publicações em periódicos do que em conferências.

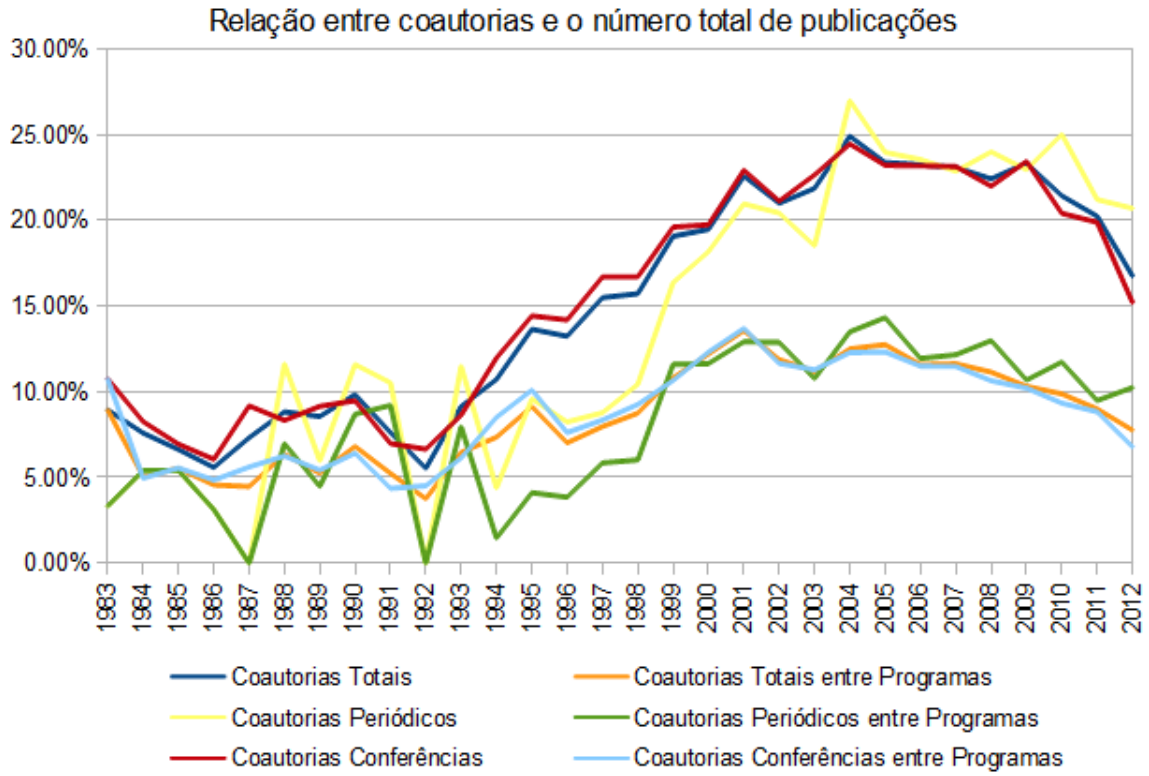

Figura 2. Evolução da relação entre coautorias e a produção total dos programas

Fonte: elaborada pelos autores a partir de dados da pesquisa.

A Figura 3 apresenta a evolução da produção acadêmica distinguindo os artigos em periódicos dos publicados em anais de conferências. É possível observar que o número de artigos publicados em periódicos tem aumentado consideravelmente em relação ao número total de artigos. De 1983 a 1992, apenas 18,31% dos artigos completos publicados pelos docentes analisados eram em periódicos. De 1993 a 2002, esta porcentagem subiu para 20,34%. De 2003 a 2012, este valor atingiu, na média, 22,24%.

As Figuras 4 a 7 apresentam redes de coautorias, 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 acadêmico de pós-graduação ao qual o docente está vinculado. O tamanho do nó é proporcional à medida de *Author Rank* do docente, considerando as relações de coautoria.

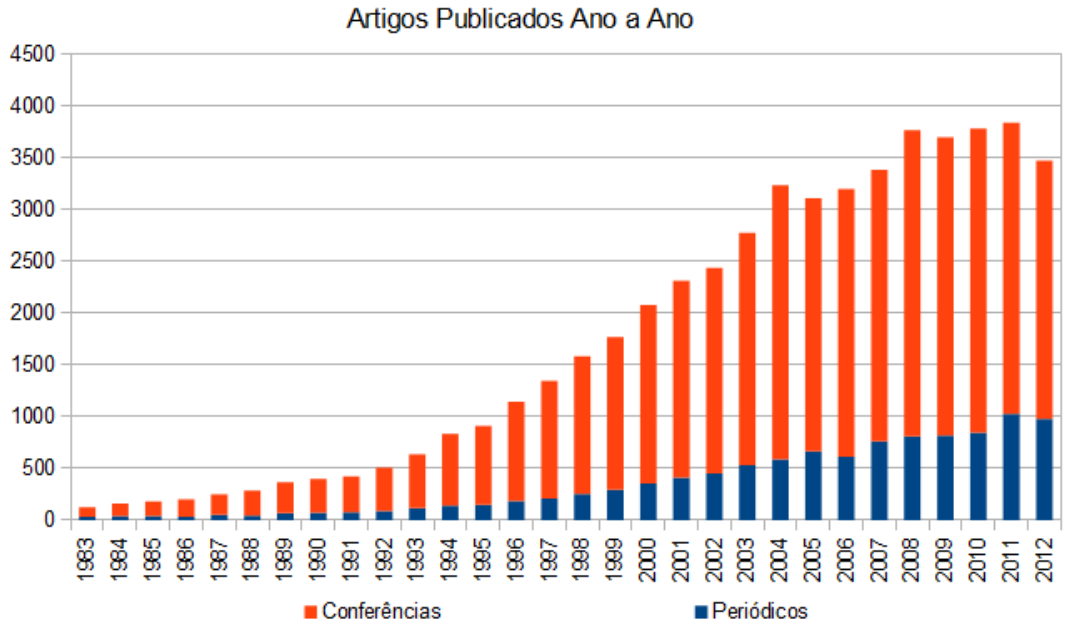

Figura 3. Evolução da produção total

Fonte: elaborada pelos autores a partir de dados da pesquisa.

As Figuras 4 e 5 contêm as redes de coautorias entre docentes nos dez triênios analisados. 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 analisado. Diversas pequenas *cliques* podem ser observadas nestas redes, formadas principalmente por integrantes de um mesmo programa (nós de uma mesma cor). Além disso, é possível observar um grande aumento nas coautorias ao longo dos triênios e aumento no número de componentes fortemente conectadas e de *cliques* nos grafos correspondentes a estas redes sociais.

A Figura 6 contém a rede de coautorias da produção anual de nove anos selecionados. Três características importantes destas redes confirmam informações observadas nas Figuras 4 e 5: (i) a grande evolução no número de coautorias; (ii) a formação de componentes fortemente conectadas e *cliques* dentro de programas; e (iii) a posição de destaque de alguns docentes na maioria dos anos considerados na análise.

A Figura 7 contém as coautorias acumuladas entre docentes. Para a análise foram utilizadas as publicações a partir de 1983. As redes de coautorias com o total acumulado dessas publicações e apresentadas nestas figuras variam dos períodos de 1983-1994 até 1983-2012. Com esta visão acumulada das coautorias é possível investigar a forte conexão entre docentes de um mesmo programa (arestas coloridas) e entre alguns docentes de diferentes programas (arestas cinzas). Em especial, ao se observar as redes da Figura 7 é possível notar que praticamente todos os nós de maior tamanho possuem uma forte ligação com ambos: docentes do mesmo programa e docentes de outros programas, correspondendo, provavelmente, às ligações entre grupos de pesquisa de diferentes programas.

Figura 4. Coautorias entre docentes: Triênios de 1983 a 1994

Fonte: elaborada pelos autores a partir de dados da pesquisa.

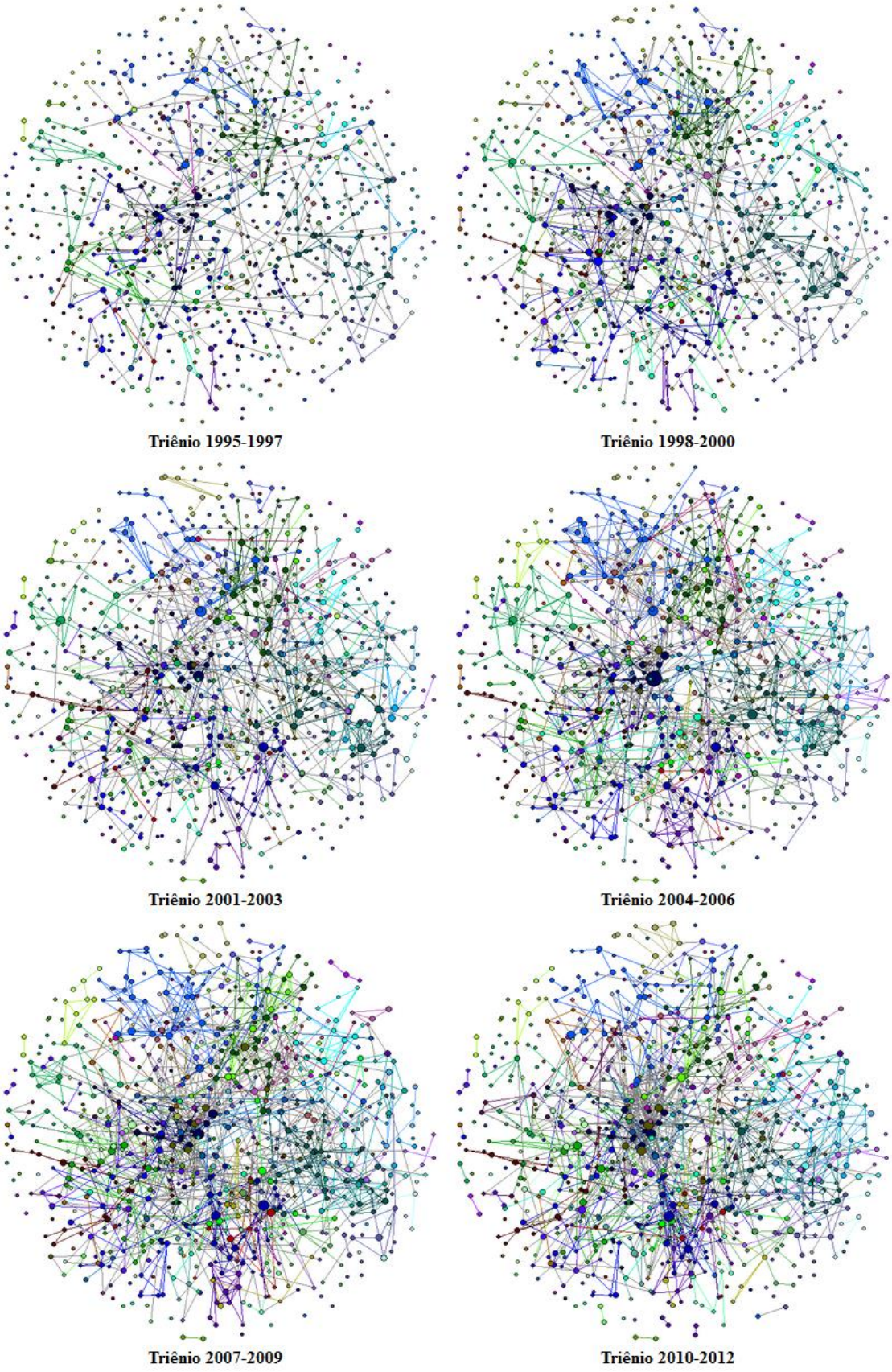

Figura 5. Coautorias entre docentes: Triênios de 1995 a 2012

Fonte: elaborada pelos autores a partir de dados da pesquisa.

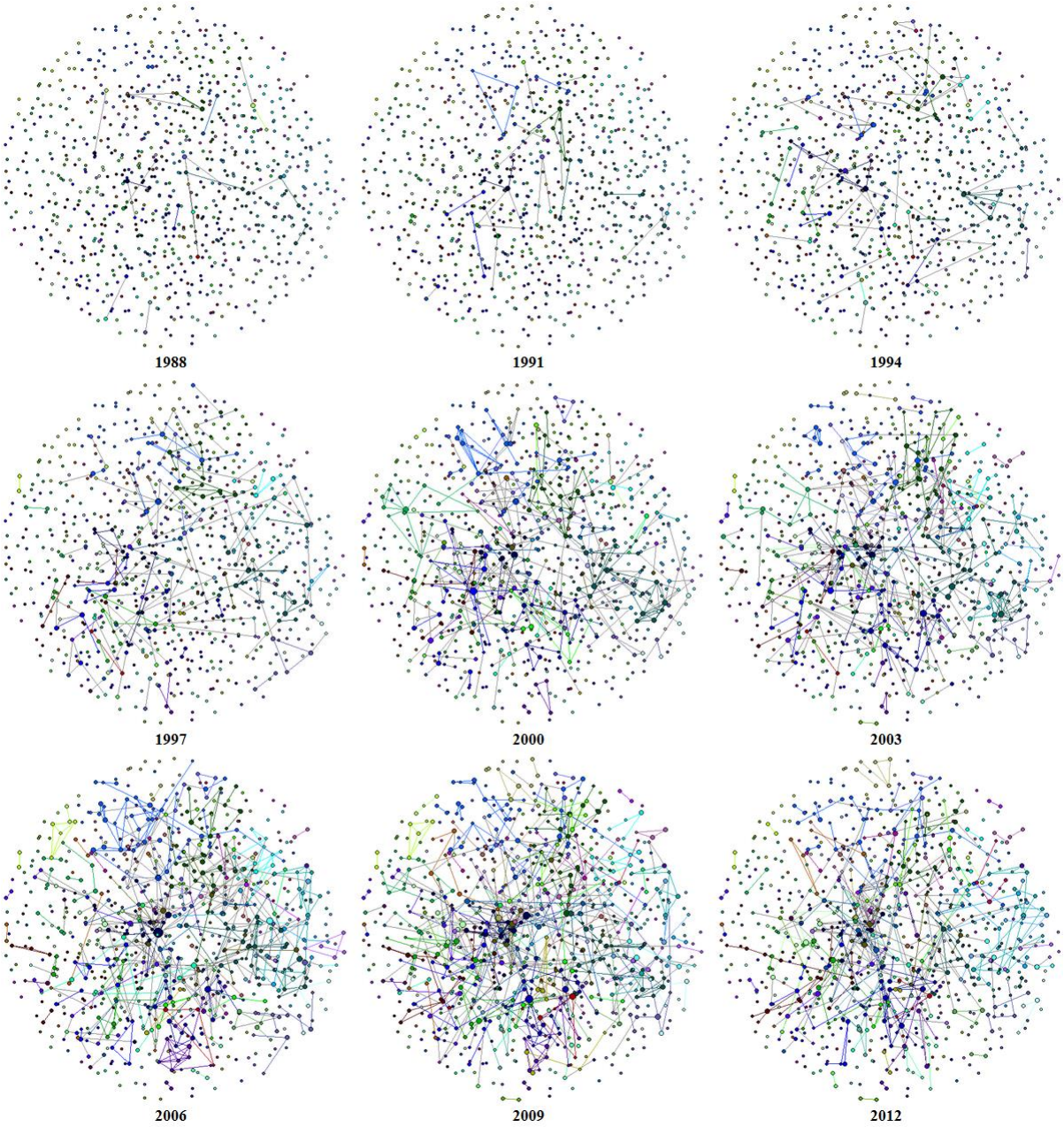

Figura 6. Coautorias entre docentes: Produção anual

Fonte: elaborada pelos autores a partir de dados da pesquisa.

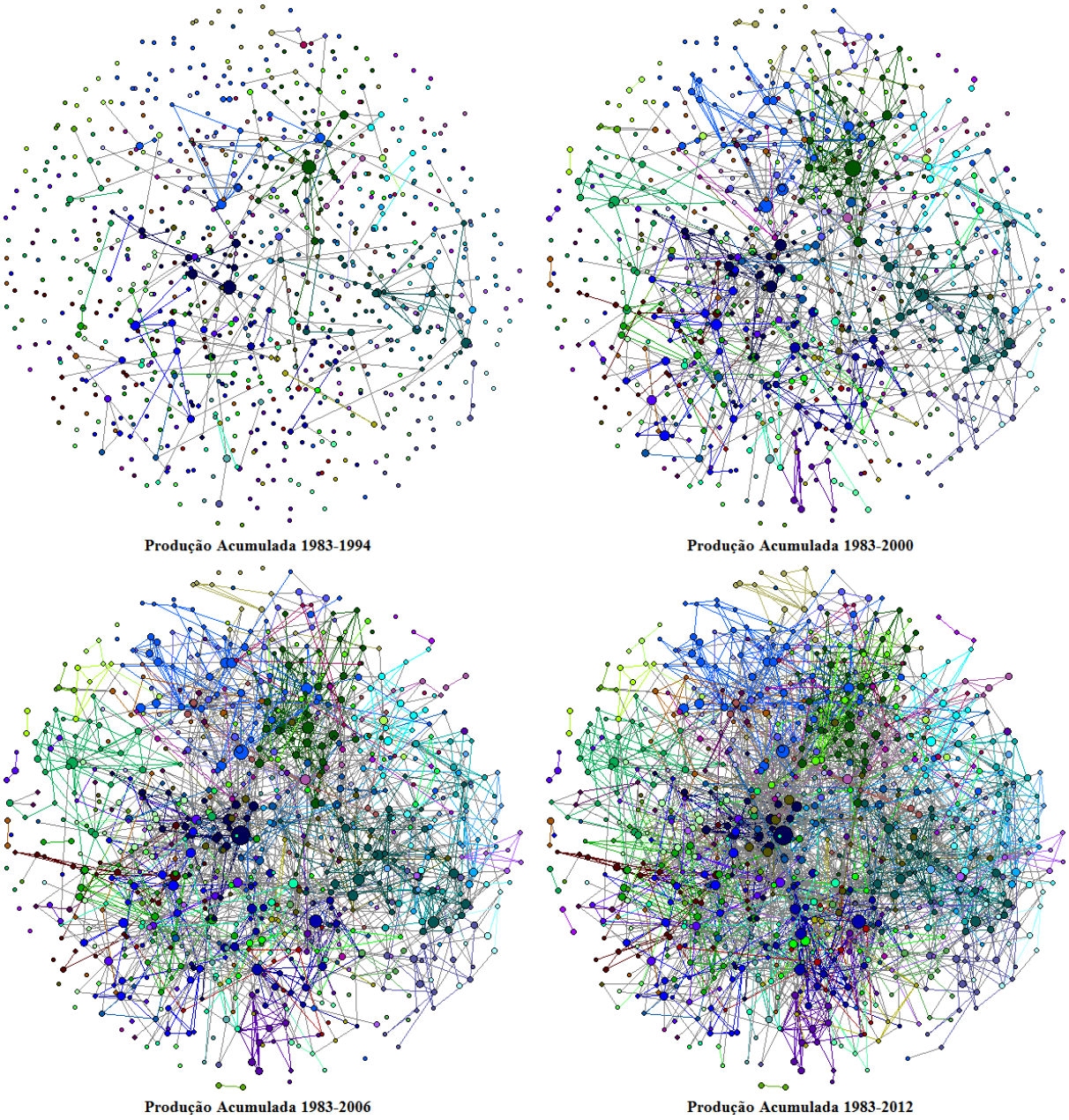

Figura 7. Coautorias entre docentes: Produção acumulada

Fonte: elaborada pelos autores a partir de dados da pesquisa.

Algumas das características das redes apresentadas nas Figuras 4 a 7 podem ser observadas na Tabela 3, que contém o número total de componentes conexas (só foram consideradas componentes com mais de um nó); o número total de nós nessas componentes; e o tamanho da componente gigante (isto é, número de nós da maior componente conexa de cada rede). Estas diferentes medidas foram normalizadas e apresentadas graficamente nas Figuras 8 e 9.

Tabela 3. Características das redes de coautoria

	Produção anual				Produção acumulada		
Ano	Número de componentes	Quantidade de nós nas componentes	Tamanho da componente gigante		Número de componentes	Quantidade de nós nas componentes	Tamanho da componente gigante
1983	3	11	6		3	11	6
1984	6	14	3		6	20	9
1985	9	19	3		10	29	9
1986	7	15	3		12	37	10
1987	15	31	3		23	63	11
1988	15	36	4		27	81	11
1989	22	51	8		31	101	21
1990	14	44	7		31	117	32
1991	19	47	5		34	134	33
1992	21	52	5		37	153	39
1993	24	69	5		37	181	45
1994	38	112	9		43	232	60
1995	52	149	9		45	283	95
1996	64	189	13		40	339	181
1997	76	230	11		40	409	224
1998	85	296	20		27	474	383
1999	79	310	65		27	530	434
2000	96	381	34		27	601	526
2001	108	413	28		25	657	585
2002	104	429	55		23	697	635
2003	95	437	50		20	728	677
2004	92	533	76		15	759	724
2005	92	514	88		12	779	751
2006	102	516	86		12	802	774
2007	86	554	297		9	817	800
2008	83	567	236		7	831	817
2009	95	597	284		5	842	832
2010	92	539	231		6	852	840
2011	88	543	236		7	858	844
2012	90	445	176		7	859	845

Fonte: elaborada pelos autores a partir de dados da pesquisa.

Na Figura 8 é interessante notar o grande salto no tamanho da componente gigante (maior componente conexa), no ano de 2007, que ocorreu devido às primeiras coautorias entre docentes, que até então só publicavam em grupos diferentes. Estas coautorias causaram a união destes grupos. Outro fator interessante é o crescimento das três medidas analisadas ao longo dos anos. Com o aumento da produção era esperado que o tamanho da componente gigante, bem como o número total de nós nas componentes, aumentasse, o que foi observado. Este tipo de aumento

costuma causar inicialmente um aumento no número de componentes (conforme observado) e uma diminuição deste número com o passar do tempo (pois os componentes começam a se fundir). Ocorreu uma redução no número de componentes, bastante sutil, depois do ano de 2001.

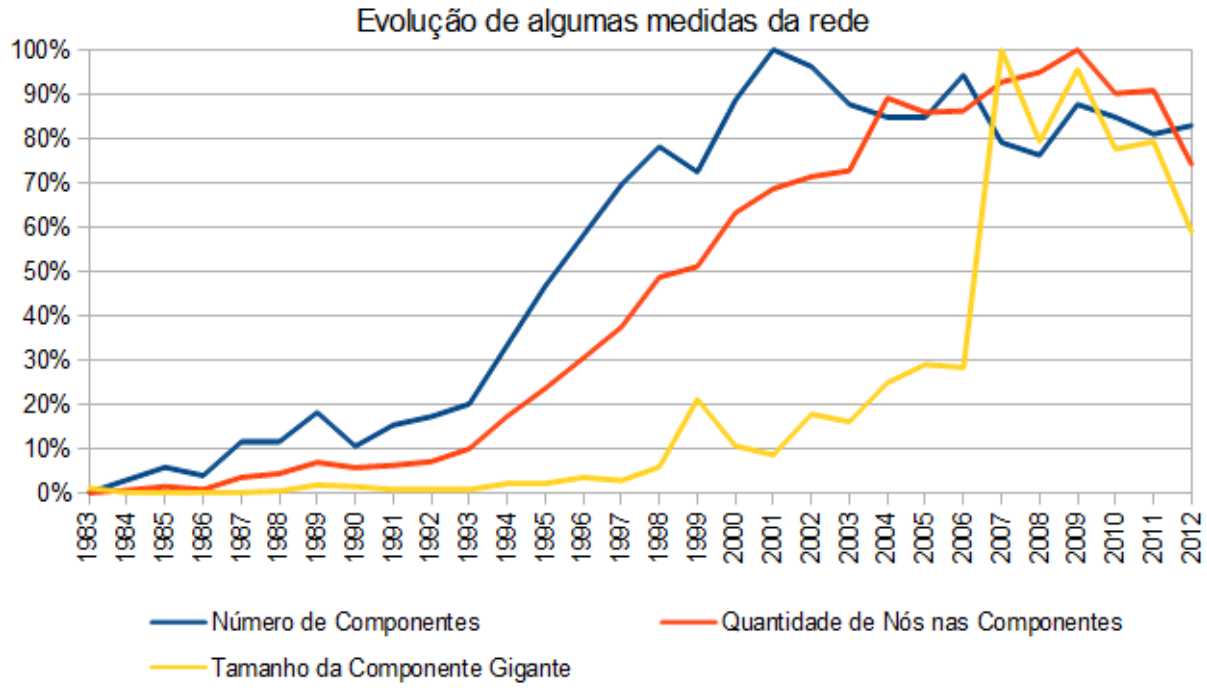

Figura 8. Evolução de algumas métricas das redes de coautoria – Produção anual

Fonte: elaborada pelos autores a partir de dados da pesquisa.

Na Figura 9 são observadas algumas métricas das redes de coautoria considerando a produção acumulada dos docentes nos trinta anos analisados. Ao contrário da Figura 8, nesta figura podemos observar claramente a redução no número de componentes conexas a partir de 1996, causada pela fusão de duas ou mais componentes. O pequeno aumento no número de componentes nos anos de 2010 e 2011 se deve ao fato de alguns docentes que, até então, não haviam participado de coautorias, terem suas primeiras coautorias neste período, criando assim novas componentes. Vale lembrar aqui que apenas componentes com mais de um docente foram computadas.

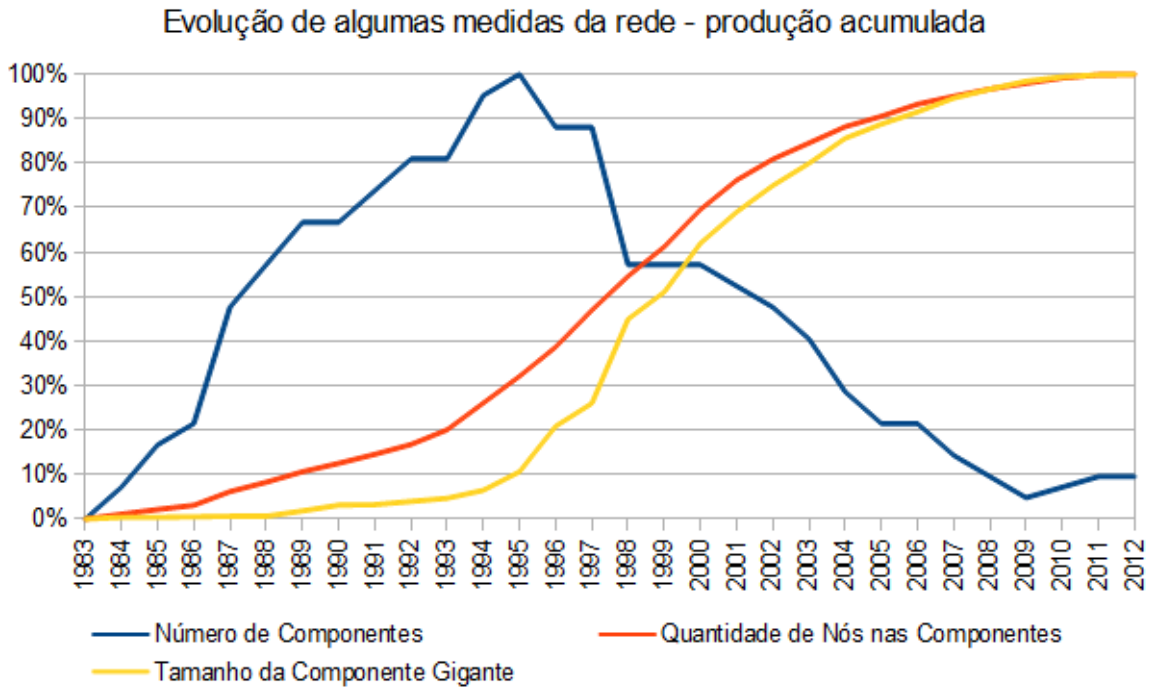

Figura 9. Evolução de algumas métricas das redes de coautoria – Produção acumulada

Fonte: elaborada pelos autores a partir de dados da pesquisa.

As Figuras 10 e 11 apresentam redes de programas de pós-graduação, onde cada nó representa um programa em Ciência da Computação avaliado pela CAPES no triênio 2007-2009. O tamanho de cada nó é proporcional ao *Author Rank* do programa, considerando as publicações em periódicos e conferências. As relações entre os nós são definidas por relações de coautoria (entre os docentes dos diferentes programas).

A Figura 10 contém as coautorias entre programas em cada um dos dez triênios analisados. 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 mostradas.

A Figura 11 contém as coautorias acumuladas entre programas nos períodos de 1983-1998 a 1983-2012. É possível verificar que as últimas redes formam uma única componente conexa. Outra característica interessante é o destaque (caracterizado pelo tamanho dos nós) que alguns programas têm em praticamente todas as redes, indicando a influência destes programas na produção científica nacional em Ciência da Computação.

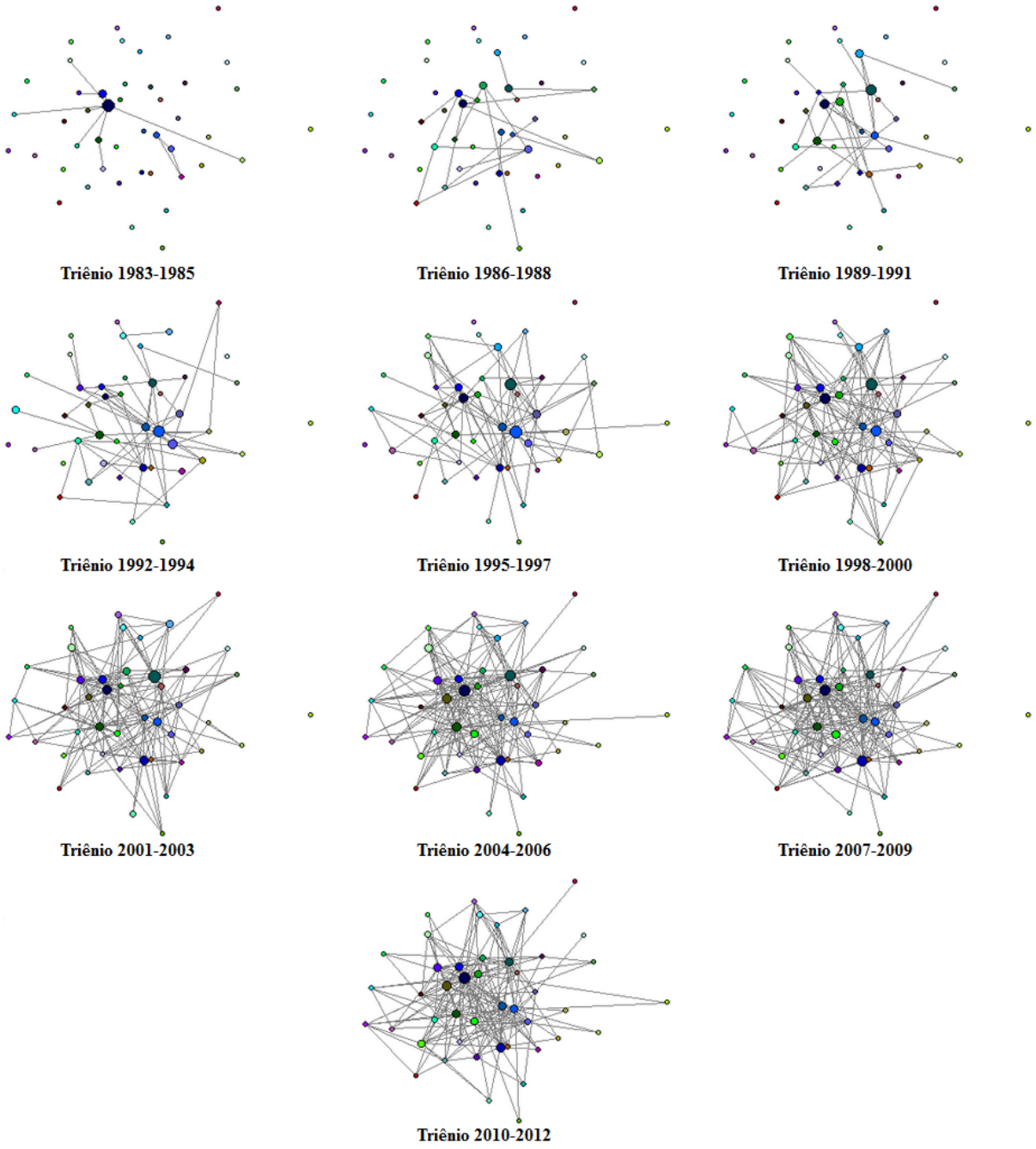

Figura 10. Coautorias entre programas

Fonte: elaborada pelos autores a partir de dados da pesquisa.

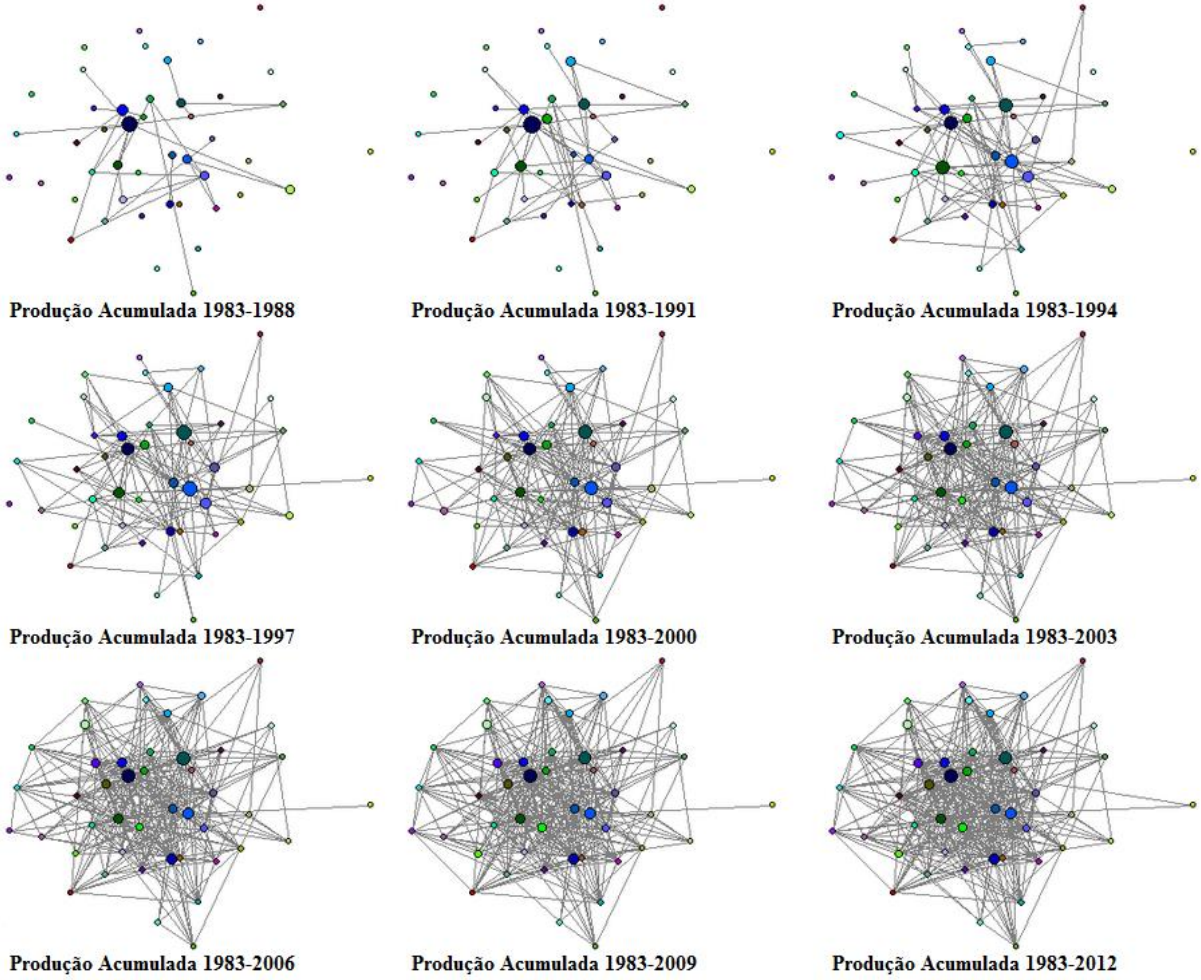

Figura 11. Coautorias entre programas – Produção acumulada

Fonte: elaborada pelos autores a partir de dados da pesquisa.

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 ou conferências.

Assim como em diversos trabalhos correlatos, neste artigo foi analisado um grupo específico de pesquisadores (aqueles que são docentes permanentes em programas de pós-graduação em Ciência da Computação no Brasil). Além da análise das redes dos docentes em um período de trinta anos, neste artigo também foram analisadas as redes formadas pelos programas de pós-graduação. Apesar de a análise de um novo grupo por si só já possuir 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 permitiu que as análises fossem realizadas de modo mais preciso considerando dois eixos importantes: pesquisadores e programas de pós-graduação.

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 de 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 de coautoria obtidas 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, detalhar a função que os principais docentes de cada programa exercem para o desenvolvimento da rede formada pelo programa, bem como as 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 pela FAPESP (Projeto Jovem Pesquisador processo 2009/10413-5 e Bolsas de Iniciação Científica processos 2011/07968-5 e 2013/06084-1), pelo CNPq (Bolsa de Iniciação Científica e Bolsa Produtividade em Pesquisa processo 304937/2010-0 e 306046/2013-0) e pelo Programa de Educação Tutorial (MEC/SESu).

REFERÊNCIAS

ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. LattesMiner: a multilingual DSL for information extraction from Lattes platform. *In: Proceedings of SPLASH'11, SPLASH'11 Workshops*, p. 85-92, New York, NY, USA. ACM, 2011. http://dx.doi.org/10.1145/2095050.2095065

BAUMES, J.; CHEN, H. C. J.; FRANCISCO, M.; GOLDBERG, M.; MAGDON-ISMAIL, M.; WALLACE, W. Visage: a virtual laboratory for simulation and analysis of social group evolution. *ACM Transactions on Autonomous and Adaptive Systems*, v. 3, n. 8, p. 1-8, 2008. http://dx.doi.org/10.1145/1380422.1380423

BERGER-WOLF, T. Y.; SAIA, J. A framework for analysis of dynamic social networks. *In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 523-528, New York, NY, USA, 2006. http://dx.doi.org/10.1145/1150402.1150462

DIGIAMPIETRI, L.; MENA-CHALCO, J. P.; PÉREZ-ALCÁZAR, J. J.; TUESTA, E. F.; DELGADO, K.; MUGNAINI, R. Minerando e caracterizando dados de currículos Lattes. *In: Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2012)*, 2012a.

DIGIAMPIETRI, L.; MENA-CHALCO, J.; SILVA, G. S.; OLIVEIRA, L.; MALHEIROS, A.; MEIRA, D. Dinâmica das relações de coautoria nos programas de pós-graduação em computação no Brasil. *In: Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2012),* 2012b.

DIGIAMPIETRI, L. A.; MENA-CHALCO, J. P.; MELO, P. O. V.; MALHEIROS, A. P.; MEIRA, D. N. O.; FRANCO, L. F.; OLIVEIRA, L. B. BraX-ray: an X-Ray of the Brazilian computer science graduate programs. *Plos One*, v. 9, p. e94541, 2014. http://dx.doi.org/10.1371/journal.pone.0094541

GUO, L.; TAN, E.; CHEN, S.; ZHANG, X.; ZHAO, Y. E. 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*, p. 369-378, New York, NY, USA. ACM, 2009. http://dx.doi.org/10.1145/1557019.1557064

HAYAT, Z.; LYONS, K. 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., 2010. http://dx.doi.org/10.1145/1923947.1923949

HORN, D. B.; FINHOLT, T. A.; BIRNHOLTZ, J. P.; MOTWANI, D.; JAYARAMAN, S. 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*, p. 582-591, New York, NY, USA. ACM, 2004. http://dx.doi.org/10.1145/1031607.1031707

KANG, H.; GETOOR, L.; SINGH, L. Visual analysis of dynamic group membership in temporal social networks. *SIGKDD Explor. Newsl.*, v. 9, n. 2, 13-21, 2007. http://dx.doi.org/10.1145/1345448.1345452

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, J. R. P. S.; DE ALENCAR, R. O.; CARDOSO, T. N.; SALLES, T. 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*, p. 427-428, New York, NY, USA. ACM, 2011. http://dx.doi.org/10.1145/1998076.1998168

LEVENSHTEIN, V. I. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, v. 10, n. 8, p. 707-710, 1966.

LIU, X.; BOLLEN, J.; NELSON, M.; de SOMPEL, H. V. Co-authorship networks in the digital library research community. *Information Processing and Management*, v. 41, n. 6, p. 1462-1480, 2005. http://dx.doi.org/10.1016/j.ipm.2005.03.012

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, n. 4, 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. IEEE, 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.

MENA-CHALCO, J. P.; DIGIAMPIETRI, L. A.; CESAR-JR, R. M.; LOPES, F. M. Brazilian bibliometric coauthorship networks. *Journal of the Association for Information Science and Technology*, v. 65, p. 1424-1445, 2014. http://dx.doi.org/10.1002/asi.23010

SHARMA, M.; URS, S. R. Network dynamics of scholarship: a social network analysis of digital library community. *In: Proceedings of the 2nd PhD workshop on Information and knowledge management*, p. 101-104, New York, NY, USA, 2008. http://dx.doi.org/10.1145/1458550.1458570

WU, B.; YE, Q.; YANG, S.; WANG, B. 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*, p. 3-10, New York, NY, USA. ACM, 2009. http://dx.doi.org/10.1145/1651274.1651277