



IDENTIFICAÇÃO E ANÁLISE DE COMUNIDADES DE COLABORAÇÃO CIENTÍFICA: ESTUDO DE CASO EM UM PROGRAMA DE PÓS-GRADUAÇÃO INTERDISCIPLINAR

Andrea Sabedra Bordin

Doutoranda em Engenharia e Gestão do Conhecimento pela Universidade Federal de Santa Catarina, Brasil.
Professora da Universidade Federal do Pampa, Brasil.
E-mail: andreabord@gmail.com

Alexandre Leopoldo Gonçalves

Doutor em Engenharia da Produção pela Universidade Federal de Santa Catarina, Brasil.
Professor da Universidade Federal de Santa Catarina, Brasil.
E-mail: a.l.goncalves@ufsc.br

João Artur de Souza

Doutor em Engenharia da Produção pela Universidade Federal de Santa Catarina, Brasil.
Professor da Universidade Federal de Santa Catarina, Brasil.
E-mail: jartur@egc.ufsc.br

Michele Andréia Borges

Doutoranda em Engenharia e Gestão do Conhecimento pela Universidade Federal de Santa Catarina, Brasil.
E-mail: micheleandborges@gmail.com

Danielly Oliveira Inomata

Doutoranda em Ciência da Informação pela Universidade Federal de Santa Catarina, Brasil.
E-mail: inomata.danielly@gmail.com

Resumo

A colaboração científica permite que esforços de pesquisa sejam unidos em prol do crescimento e da qualidade da ciência. O objetivo deste artigo foi identificar e analisar as comunidades de colaboração científica em um Programa de Pós-Graduação interdisciplinar brasileiro. Utilizou-se o método de Análise de Redes Sociais (ARS) para detecção de comunidades e para cada comunidade foram extraídos e analisados os termos mais frequentes a partir do título das produções intelectuais. Foram encontradas nove comunidades de colaboração com arranjos distintos e cada comunidade apresentou um corpo de termos que permeiam as três áreas de concentração do programa. Os resultados encontrados demonstraram que existe um equilíbrio entre as áreas de concentração do programa. Entretanto, comunidades isoladas devem ser evitadas e um trabalho mais interdisciplinar entre algumas áreas de concentração deve ser fomentado.

Palavras-chave: Rede de colaboração científica. Rede de coautoria. Análise de rede social. Detecção de comunidades. Extração de informação.

**IDENTIFICATION AND ANALYSIS OF SCIENTIFIC COLLABORATION'S COMMUNITIES:
A CASE STUDY IN A PROGRAM OF GRADUATE INTERDISCIPLINARY**

Abstract

The scientific collaboration enables research efforts are joined in support of growth and quality of science. The purpose of this article is to identify and analyze the communities of scientific collaboration in a Brazilian Interdisciplinary Program Postgraduate. We used the method of Social Network Analysis (SNA) for community detection. For each community were extracted and analyzed the most frequent terms from the title of intellectual productions. Nine communities of collaboration were found with different arrangements and each community has a body of terms that permeate the three focus areas of the program. The results show that there is a balance between the focus areas of the program. However, isolated communities must be avoided and a more interdisciplinary work between some areas of concentration should be encouraged.

Keywords: *Scientific collaboration network. Co-authorship network. Social network analysis. Community detection. Information extraction.*

1 INTRODUÇÃO

A colaboração científica, tanto formal como informal, é uma das características que definem a ciência moderna (MILOJEVIC, 2010). Conforme Katz e Martin (1997) são inúmeros os benefícios da comunicação científica, seja relacionado ao compartilhamento de conhecimentos, habilidades e técnicas inerentes a essa prática; a transferência de conhecimentos, principalmente o conhecimento tácito; seja pelo fato da colaboração trazer novos enfoques com relação a fertilização de ideias, que por sua vez pode gerar novas perspectivas em relação a pesquisadores que trabalham individualmente.

A colaboração é capaz de promover a atividade de pesquisa, a produtividade e o impacto, portanto, deve ser encorajada e suportada por meio de mecanismos de gestão de pesquisa e políticas de ciência (GLÄNZEL; SCHUBERT, 2004).

Embora o conceito colaboração científica seja de difícil definição Katz e Martin (1997), Hara et al. (2003) sugerem que a colaboração pressupõe ao menos dois elementos: i) trabalhar conjuntamente por um objetivo em comum e ii) compartilhar conhecimento. Partindo dos elementos que a colaboração pressupõe, pode-se dizer que a colaboração científica acontece quando o trabalho e o compartilhamento ocorrem em prol de um conhecimento científico.

Segundo Abbasi, Altmann e Hwang (2010), no meio acadêmico a coautoria é o mais visível e acessível indicador de colaboração científica e tem sido usado frequentemente para medir a atividade colaborativa especialmente em estudos bibliométricos (GLÄNZEL; SCHUBERT, 2004) e estudos de análise de rede social (MILOJEVIC, 2010; NEWMAN, 2001).

Uma rede de colaboração científica pode ser vista como um tipo de rede social onde os colaboradores ou pesquisadores são os nodos ou atores e os relacionamentos de colaboração são as arestas que ligam os nodos. Ao analisar uma rede de colaboração científica sob esse aspecto é possível aplicar uma série de medidas/métricas de análise de rede tais como: densidade, diâmetro, centralidade de atores, detecção de comunidades de atores dentro da rede, etc. Redes de colaboração científica frequentemente são vistas e encontradas na literatura como sinônimos de redes de coautoria (VANZ; STUMPF, 2010).

A detecção de comunidades numa rede de colaboração científica ou de coautoria permite identificar os grupos cujos laços de colaboração são mais densos (frequente) entre os atores do grupo e mais fracos (esparcos) entre os atores externos ao grupo. Ao detectar tais comunidades é possível saber quantos grupos densos de colaboração existem, qual o tamanho

de cada grupo, quem são os atores que formam os grupos e inferir outras informações como as áreas de pesquisa de cada grupo. Para cada comunidade detectada, a qual pode ser considerada uma sub-rede, é possível aplicar novamente métricas de análise de rede social para descobrir, por exemplo, quais os atores mais centralizados, efetuando assim análises mais refinadas (MARTINS, 2012).

A produção intelectual (artigos em periódicos, artigos em anais de evento, livro e capítulo de livro) do corpo docente é um requisito com significativo grau de impacto nas avaliações dos programas de pós-graduação *stricto sensu* (mestrado e doutorado) no Brasil (ROSAS, 2013). Deste modo, a colaboração científica entre docentes e discentes é de suma importância para aumentar o impacto das produções intelectuais (WEISZ; ROCO, 1996; BALANCIERI et al., 2005).

Quando se trata de programas de pós-graduação interdisciplinar, a colaboração científica torna-se um importante mecanismo para tornar os diálogos mais amplos, uma vez que as produções intelectuais dos programas de pós-graduação interdisciplinares devem visar um saber cujos limites ultrapassam as fronteiras disciplinares (CAPES, 2013).

Deste modo, investigar a questão de “como se formam as comunidades científicas em um programa de pós-graduação interdisciplinar?” permite avançar na compreensão da estrutura desse tipo de rede, nas suas características e particularidades, que pode resultar em esforços para melhorar a qualidade das produções intelectuais na área interdisciplinar, bem como, em medidas para fortalecer e ampliar a sua rede de colaboração científica.

Na tentativa de responder a indagação acima, este artigo se propõe a identificar e analisar as comunidades de colaboração científica em um Programa de Pós-Graduação interdisciplinar brasileiro, por meio de métodos de análise de redes sociais e da análise dos termos presentes nas publicações dos autores das comunidades, buscando conhecer a gênese da formação de comunidades científicas desse programa de pós-graduação, que reúnem tanto um corpo docente como discente com formações em áreas distintas e transversais.

O artigo está organizado da seguinte forma: referencial teórico a respeito da interdisciplinaridade e da rede de colaboração científica; procedimentos metodológicos realizados para alcançar o objetivo proposto neste artigo; resultados obtidos; discussão dos resultados; e considerações finais.

2 REFERENCIAL TEÓRICO

Nesta seção serão abordados dois tópicos basilares para a construção deste artigo: o conceito de interdisciplinaridade no contexto da pós-graduação e as redes de colaboração científica.

2.1 Interdisciplinaridade

Para entendermos o conceito de interdisciplinaridade temos que entender o conceito de disciplina. Uma definição própria de disciplina, de acordo com Pineau (2005), é o conjunto específico de conhecimentos que tem suas características próprias no plano de ensino, da formação, dos mecanismos, dos métodos e matérias.

Já a interdisciplinaridade é algo que ultrapassa as fronteiras do saber disciplinar, isto é, avança no sentido de uma combinação, de uma convergência, de uma complementaridade (POMBO, 2008).

Segundo definição da CAPES (2013, p. 12):

A interdisciplinaridade pressupõe uma forma de produção do conhecimento que implica trocas teóricas e metodológicas, geração de novos conceitos e

metodologias e graus crescentes de intersubjetividade, visando a atender a natureza múltipla de fenômenos complexos. Entende-se por Interdisciplinaridade a convergência de duas ou mais áreas do conhecimento, não pertencentes à mesma classe, que contribua para o avanço das fronteiras da ciência e tecnologia, transfira métodos de uma área para outra, gerando novos conhecimentos ou disciplinas e faça surgir um novo profissional com um perfil distinto dos existentes, com formação básica sólida e integradora.

A área Interdisciplinar no contexto da pós-graduação é introduzida decorrente da necessidade de se dar conta de novos problemas que emergem no mundo contemporâneo, de diferentes naturezas e com variados níveis de complexidade, que requerem diálogo entre disciplinas próximas, dentro da mesma área do conhecimento, mas também entre disciplinas de áreas diferentes, bem como entre saberes disciplinar e não disciplinar (CAPES, 2013).

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), que desempenha papel fundamental na expansão, consolidação e avaliação da pós-graduação *stricto sensu* (mestrado e doutorado) no Brasil, relata que nos últimos anos a área interdisciplinar vem apresentando a maior taxa de crescimento. Dentre os fatores que condicionam o crescimento de programas de pós-graduação interdisciplinar no Brasil, um deles está associado a tendência mundial de aumento de grupos de pesquisa e programas acadêmicos com foco em questões complexas, induzindo assim, o fomento de cursos em áreas inovadoras e interdisciplinares (CAPES, 2013).

Para a consolidação dos Programas de Pós-graduação alguns requisitos e orientações são determinados e indicados pela CAPES. A produção intelectual (artigos em periódicos, artigos em anais de evento, livro e capítulo de livros) do corpo docente é um desses requisitos, com significativo grau de impacto nas avaliações dos programas de pós-graduação (CAPES, 2013).

Em programas de pós-graduação interdisciplinar a produção intelectual do corpo docente deve ser coerente com a área de concentração e linha de pesquisa do docente. Além disso, as produções devem visar um saber cujos limites ultrapassam as fronteiras disciplinares. Deste modo, a colaboração científica é de suma importância para romper as barreiras disciplinares bem como aumentar o impacto das produções intelectuais (CAPES, 2013).

2.2 Redes de Colaboração Científica

Embora seja descrito na literatura que as comunidades científicas se formam por um objetivo em comum, e pelo compartilhamento do conhecimento (KATZ; MARTIN, 1997; HARA et al., 2003), precisar as causas da formação de redes de colaboração científica ainda é uma questão em aberto (MARTINS; FERREIRA, 2013). É válido ressaltar que com a análise das redes de colaboração científica é possível identificar “colégios invisíveis”, ou seja, pesquisadores que realizam trabalho conjuntamente por compartilharem o mesmo objeto de pesquisa, porém não estão fisicamente próximos, não trabalham na mesma instituição e o idioma e a nacionalidade são distintos (FERRAZ; QUONIAM; ALVARES, 2014).

Martins e Ferreira (2013, p. 187) destacam que na área de bibliometria e cientometria emergem, nos últimos anos, trabalhos que apontam a análise de redes sociais como método de estudo na colaboração científica, e com ela o uso de métricas “para explicar possíveis causas que determinam o desempenho de pesquisadores e instituições em seus modos de produção científica”.

A análise de rede social estuda as relações sociais sob o aspecto de uma rede. Segundo a definição da INSNA (*Internacional Network for Social Network Analysis*), a análise de redes

sociais está focada em descobrir os padrões de interações das unidades de análise (pessoas, organizações, países). Uma rede pode ser representada por um grafo $G = (V, E)$ formado por Vértices (V) e Arestas (E). Cada vértice ou nodo representa um ator e cada aresta representa a relação existente entre dois atores integrantes da rede. No caso de uma rede de coautoria, os vértices representam os autores e as arestas representam as produções bibliográficas produzidas em parceria. De modo geral, uma rede pode ser direcionada ou não direcionada e as arestas podem ser valoradas ou não. No caso da rede de coautoria, a rede é não direcionada e valorada, porque a relação de coautoria é assíncrona e a valoração se dá em função do número de publicações em conjunto.

Uma maneira de entender uma rede e seus atores é avaliar a localização desses atores na rede em termos de suas posições estratégicas. O conceito de centralidade e as medidas associadas a esse conceito levam em consideração as diferentes formas com que um ator interage e se comunica com o restante da rede. Os atores mais importantes ou mais proeminentes estão normalmente localizados em posições estratégicas dentro da rede (WASSERMAN; FAUST, 1994).

Encontra-se difundido na literatura os termos conectores ou *hub*. Os conectores ou *hubs* são nodos que contém um número anormalmente grande de links, ou seja, são nodos altamente conectados. Em outras palavras, enquanto alguns indivíduos têm poucas conexões, outros têm uma gama enorme de conexões, esse último é considerado *hub*, ou conector potencial (BARABÁSI, 2009).

Para quantificar a importância de um ator em uma rede social, várias medidas de centralidade foram propostas ao longo dos anos (FREEMAN, 1979; SCOTT, 2000). A centralidade de grau é uma medida de centralidade local determinada pelo grau do nodo na rede, sendo o grau do nodo determinado pelo número de arestas ou ligações desse nodo. No caso de redes valoradas, o valor ou peso da aresta é levado em conta no cálculo do grau.

Em uma rede de coautoria, esse grau indica o número de autores da rede que publicaram em parceria com um determinado autor. O grau de centralidade varia de 0 (indicando que o autor publicou sozinho) até $n-1$ (indicando que o autor publicou com todos os demais autores da rede excluindo a si próprio), no qual n é o número total de autores da rede.

Segundo Newman (2006), uma questão que tem recebido considerável atenção por parte da comunidade de sistemas de rede é a detecção e caracterização da estrutura de comunidades em redes. Comunidades são grupos densamente conectados através de vértices e com ligações esparsas entre os grupos. A capacidade de detectar tais grupos pode ter uma importância prática significativa. Por exemplo, grupos dentro de redes sociais podem corresponder a unidades sociais ou comunidades (GIRVAN; NEWMAN, 2002).

Trabalhos anteriores sobre os métodos de detecção de comunidades ou descoberta de grupos em redes dividem-se em duas linhas de pesquisa, ambas com uma longa história. A primeira atende pelo nome de particionamento de grafo e tem sido estudada particularmente em ciência da computação e áreas afins, com aplicações em computação paralela e design de circuitos integrados, entre outras áreas (HENDRICKSON; KOLDA, 2000). A segunda, conhecida como modelagem de bloco (*block modelling*), agrupamento hierárquico (*hierarchical clustering*), ou detecção de estrutura de comunidade (*community detection*), tem sido estudada por sociólogos e, mais recentemente, por físicos, biólogos, e matemáticos aplicados, com aplicações especialmente em redes sociais e biológicas (WASSERMAN; FAUST, 1994).

Existem diferenças importantes entre os objetivos das duas linhas de pesquisa. Um problema típico estudado em particionamento de grafo é a divisão de um conjunto de tarefas entre os processadores de um computador, de forma a minimizar a quantidade necessária de comunicação entre os mesmos. Nessa abordagem se sabe de antemão o número e o tamanho dos grupos em que a rede será dividida, pois já se sabe o número de processadores existentes.

O objetivo dessa abordagem é encontrar a melhor divisão da rede independentemente se uma boa divisão existe (NEWMAN, 2006).

A detecção de estrutura de comunidade, ao contrário, é uma técnica de análise de dados utilizada para destacar a estrutura de um conjunto de dados de uma rede de larga escala, por exemplo, uma rede social. Métodos de estrutura da comunidade normalmente assumem que a rede de interesse divide-se naturalmente em subgrupos e o trabalho do pesquisador é encontrar esses grupos. O número e tamanho dos grupos são determinados pela própria rede e não pelo experimentador (NEWMAN, 2006).

3 PROCEDIMENTOS METODOLÓGICOS

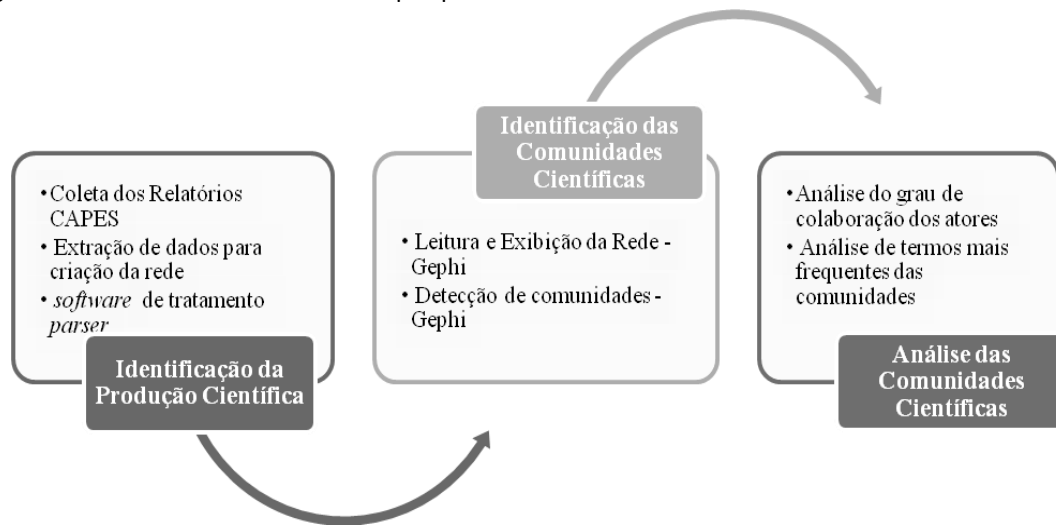
Os dados utilizados neste estudo referem-se à produção bibliográfica produzida por docentes, discentes e participantes externos ao Programa de Pós-graduação em Engenharia e Gestão do Conhecimento (PPGEGC) da Universidade Federal de Santa Catarina (UFSC).

O PPGEGC é um programa de natureza interdisciplinar que foi criado em meados de 2004 e no período de 2005 a 2012 produziu 2182 publicações informadas ao sistema Coleta de Dados CAPES. O Coleta de Dados CAPES é um sistema informatizado desenvolvido com o objetivo de coletar informações dos cursos de mestrado, doutorado e mestrado profissional integrantes do Sistema Nacional de Pós-Graduação. Para cada ano base de produção coletada foram gerados, a partir desse próprio sistema, relatórios da produção bibliográfica do tipo trabalhos em anais, artigos em periódicos, livros ou capítulos de livros. Ao total foram 2182 produções bibliográficas divididas em anais, artigos, livros e capítulos produzidos por 1488 autores diferentes. Atualmente o sistema de coleta de dados da pós-graduação é a Plataforma Sucupira.

O PPGEGC desde sua concepção tem caracterizado seu foco de pesquisa e formação no conhecimento e nos processos que lhe tornam fator gerador de valor na sociedade contemporânea. Sua estruturação em áreas de concentração é justamente decorrente da atribuição de missões inter-relacionadas aos processos de codificação/formalização (área de Engenharia do Conhecimento); planejamento e gerência (área de Gestão do Conhecimento); e difusão, comunicação e compartilhamento do conhecimento (área de Mídia e Conhecimento). O PPGEGC, portanto, considera que seu objeto de pesquisa e formação é essencialmente interdisciplinar (PACHECO; TOSTA; FREIRE, 2010).

Em seguida, serão detalhados os procedimentos adotados para a pesquisa, os quais seguiram, conforme mostra a Figura 1, basicamente três etapas: 1) Identificação da Produção Científica; 2) Identificação das Comunidades Científicas; e, 3) Análise das Comunidades Científicas.

Figura 1 – Procedimentos adotados na pesquisa



Fonte: Elaborado pelos autores

Na etapa de identificação da produção científica, primeiramente foram coletados os relatórios de produção bibliográfica produzidos pelo sistema Coleta de Dados da CAPES, o qual permite a emissão de relatórios no formato PDF por tipo de produção e ano. A partir dos relatórios foram extraídos os dados necessários para a criação da rede de colaboração ou coautoria e posterior identificação das comunidades. Para a extração de dados textuais foi desenvolvido um *software*¹ (do tipo *parser*) que extraiu dos relatórios dados de autores e relações de coautoria necessários para a representação da rede. Este software faz uma varredura em arquivos no formato PDF identificando padrões de texto, tais como “Título” e “Autor”. A partir dessa identificação são extraídos os dados associados a cada padrão.

Na segunda etapa, para identificar as comunidades de colaboração científica do PPGEGC foi utilizado o software de visualização e análise exploratória de dados de rede Gephi (BASTIAN; HEYMANN; JACOMY, 2009). O software Gephi deve ser alimentado com os arquivos contendo a relação de nodos ou autores da rede e suas relações e pesos. A partir disso, ele exibe graficamente a rede e possibilita a aplicação de diversas métricas de análise de rede. A detecção de comunidades foi feita por um método de detecção de comunidades em redes que utiliza o algoritmo de detecção de comunidades de Blondel et al. (2008) presente em um módulo do Gephi.

O algoritmo de detecção de comunidades de Blondel et al. (2008) é baseado em otimização de modularidade. O algoritmo é dividido em duas fases, as quais são repetidas iterativamente. Suponha que se comece com uma rede valorada de N nodos. Primeiro, atribui-se uma comunidade diferente para cada nodo da rede. Assim, nessa partição inicial há tantas comunidades quantos nodos. Então, para cada nodo i consideram-se os j vizinhos de i e avalia-se o ganho de modularidade que ocorreria através da remoção de i de sua comunidade para a comunidade de j . O nodo i é então colocado na comunidade para a qual este ganho é maximizado, mas apenas se este ganho é positivo. Se nenhum ganho positivo é possível, i fica em sua comunidade de origem. Este processo é aplicado repetidamente e sequencialmente para todos os nodos até que nenhuma melhoria possa ser alcançada e a primeira fase é então completada. A segunda fase do algoritmo consiste na construção de uma nova rede cujos

¹ *Software* não comercializável, desenvolvido por uma das autoras deste trabalho, para fins de extração de dados.

nodos agora são as comunidades encontradas durante a primeira fase. Para fazer isso, os valores (pesos) das ligações entre os novos nodos são dados pela soma dos valores das ligações entre os nodos nas correspondentes duas comunidades. As ligações entre os nodos de uma mesma comunidade levam à *auto-loops* para esta comunidade na nova rede. Uma vez que esta segunda fase é concluída é possível reaplicar a primeira fase do algoritmo para a rede resultante.

Na terceira etapa, para cada comunidade detectada foram feitas análises baseadas na centralidade de grau dos atores, com o objetivo de descobrir os autores que mais colaboram, e na extração dos termos mais frequentes dos títulos das produções dos atores, com o objetivo de descobrir as áreas de concentração de pesquisa da comunidade. Por termo ou entidade, entende-se uma ou mais palavras com significado bem definido.

Em uma rede de coautoria, as arestas ou relações são valoradas ou possuem um peso, dado pelo número de publicações feitas em conjunto. A centralidade de grau de atores desse tipo de rede é dada pela soma dos pesos das arestas.

Para a extração dos termos mais frequentes dos títulos das produções foi desenvolvido um *software*² que utiliza uma técnica de extração de entidades (termos compostos por uma ou mais palavras). A extração de entidades é uma tarefa de extração de informação (FELDMAN; SANGER, 2007), cujo objetivo geral é extrair informações estruturadas e úteis a partir de documentos não estruturados ou semiestruturados legíveis por máquina. O *software* extrator de termos utiliza uma abordagem de extração baseado no modelo *n-gram*. Modelos *n-gram* exploram a ordem das palavras em uma sentença e podem ser imaginados como se fosse colocada uma pequena janela sobre uma sentença ou texto, no qual somente *n* palavras são visíveis ao mesmo tempo (KOK; BROUWER, 2011). Para essa implementação foi utilizada a API *Lucene* e o analisador *ShingleAnalyzerWrapper*.

4 RESULTADOS

Na rede de colaboração do PPGEHC foram encontradas nove comunidades. A Tabela 1 mostra a relação das comunidades detectadas em ordem decrescente de número de nodos da comunidade, juntamente com o percentual da comunidade em relação à rede completa.

Tabela 1 – Comunidades detectadas na rede de colaboração do PPGEHC no período 2005-2012

N. da Comunidade	N. Nodos/autores	Percentual em relação à rede
1	358	24,1
2	324	21,8
3	263	17,7
4	216	14,5
5	197	13,2
6	115	7,7
7	11	0,7
8	3	0,2
9	1	0,1

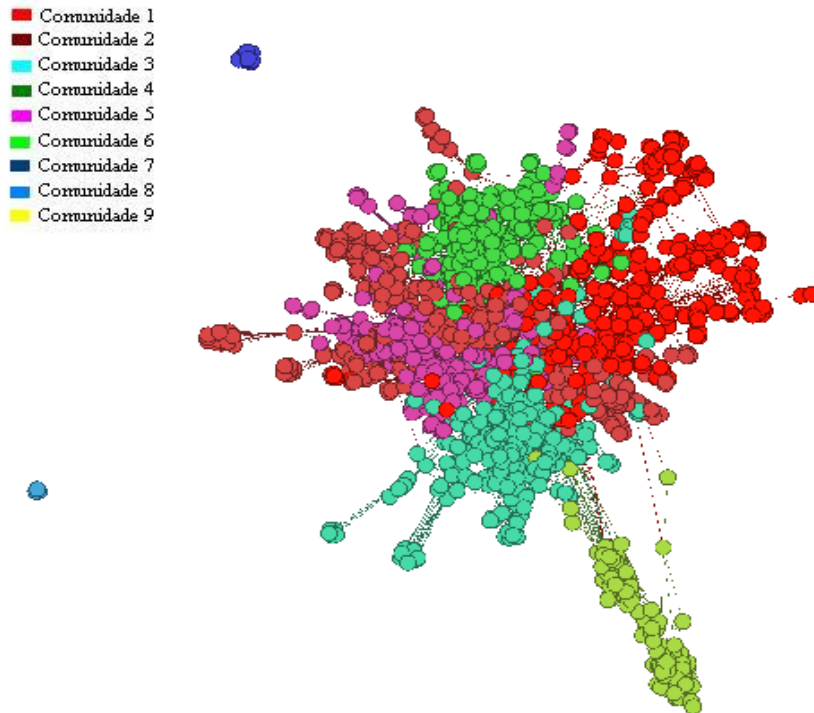
Fonte: Dados da pesquisa, 2013

A Tabela 1 evidencia a representatividade das comunidades na rede. A Comunidade 1 é a mais representativa, com 24,1% de toda a rede, seguida da Comunidade 2, com 21,8%.

² *Software* não comercializável, desenvolvido por uma das autoras deste trabalho, para fins de extração de entidades.

Com relação à estrutura da rede, a Figura 2 exibe o quanto as comunidades estão sobrepostas e quais estão isoladas. Observa-se que existem seis comunidades (sub-redes) onde todos os nodos estão conectados por algum caminho, isso significa que a rede é densa. E três comunidades menores, que estão isoladas das demais, representadas pelas Comunidades 7, 8 e 9.

Figura 2 – Densidade das comunidades de colaboração do PPGEGC no período 2005-2012



Fonte: Elaborado pelos autores

A seguir serão apresentados, para cada comunidade detectada, resultados como o número de publicações, os autores que mais colaboraram conjuntamente com outros autores e sua respectiva centralidade de grau e área de concentração. Além disso, serão apresentados os termos mais frequentes encontrados nos títulos das publicações dos autores de cada comunidade:

- Comunidade 1

Na comunidade 1 estão 358 autores que representam 24,1% de toda a rede e 564 produções bibliográficas. Os cinco autores que mais colaboram, ou com maior centralidade de grau, são P.M.S. (387), F.A.F. (240), G.J.V.R. (165), E.J.S.S. (91) e L.A.G. (86). Desses autores alguns foram e outros são docentes do PPGEGC, sendo os quatro primeiros da área de concentração Gestão do Conhecimento e o último da área Mídia e Conhecimento. A comunidade conta com mais sete docentes, totalizando 12 pesquisadores (nove da área Gestão do Conhecimento, dois da área Mídia e Conhecimento e um da área Engenharia do Conhecimento). A Figura 3 exibe os termos mais frequentes encontrados nesta comunidade.

Figura 3 – Termos mais frequentes nas publicações da Comunidade 1



Fonte: Elaborado pelos autores

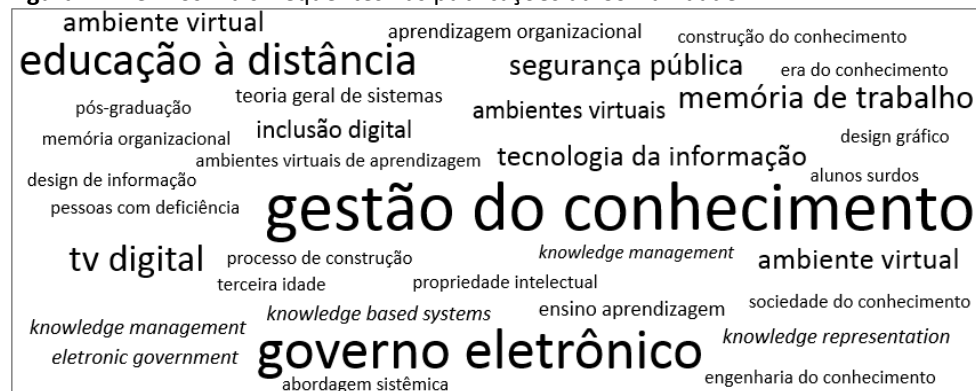
O termo Gestão do Conhecimento é o termo mais recorrente nesta comunidade. Ele aparece 60 vezes. O segundo termo mais frequente, Desenvolvimento de Produtos, tem frequência igual a 17. Entre o primeiro e segundo termo há um intervalo de frequência em 43 vezes. Os demais termos decrescem suavemente: Tecnologia da Informação (16), Processo de Desenvolvimento de Produtos (13), *Product development* (11).

Observa-se que nesta comunidade os *hubs* (os autores que mais colaboram) são predominantemente da área Gestão do Conhecimento e os termos dos títulos das produções intelectuais são aderentes a esta área de concentração.

- Comunidade 2

Na Comunidade 2 existem 324 autores que representam 21,8% de toda a rede com 652 produções. Os seis autores com maior centralidade de grau são F.A.P.F. (428), H.C.H. (216), R.P.L.S. (173), A.J.R. (170), J.B.M.A. (154) e S.S.L.F. (121). As áreas de concentração dos referidos autores, alguns ex-docentes e outros docentes do programa, estão intercaladas nesse *ranking*, sendo três autores da área Mídia e Conhecimento e os outros três da área Engenharia do Conhecimento. Nesta comunidade foram encontrados mais quatro docentes em outras posições, totalizando 10 pesquisadores (cinco da área Mídia e Conhecimento, três da área Engenharia do Conhecimento e dois da área Gestão do Conhecimento). A Figura 4 exibe os termos mais frequentes encontrados nesta comunidade.

Figura 4 – Termos mais frequentes nas publicações da Comunidade 2



Fonte: Elaborado pelos autores

Assim como na Comunidade 1, o termo mais frequente nesta comunidade é Gestão do Conhecimento (frequência igual a 40). Em segundo, encontra-se o termo Governo Eletrônico (28), objeto de estudo da área Engenharia do Conhecimento, e em terceiro o termo Educação a Distância (26), objeto de estudo da área Mídia e Conhecimento do PPGEGC.

Apesar do termo Gestão do Conhecimento ter uma frequência alta nas produções intelectuais desta comunidade, as áreas Engenharia do Conhecimento e Mídia e Conhecimento estão bem representadas por termos que são foco de pesquisa dessas duas áreas. Esta característica da comunidade vai ao encontro com as áreas dos autores *hubs*.

- Comunidade 3

A Comunidade 3 é composta por 263 autores com 436 produções que representam 17,7% de toda a rede. Os quatro autores com maior centralidade de grau são C. J. C. A. C. (187), R. C. P. (179), J. L. T. (177) e F. A. O. G. (167). Esses quatro autores são docentes do programa, sendo o primeiro da área de concentração Gestão do Conhecimento e os demais da área Engenharia do Conhecimento. A comunidade possui mais cinco docentes, totalizando nove pesquisadores (oito da área Engenharia do Conhecimento e um da área Gestão do Conhecimento). A Figura 5 exibe os termos mais frequentes encontrados nesta comunidade.

Figura 5 – Termos mais frequentes nas publicações da Comunidade 3



Fonte: elaborado pelos autores.

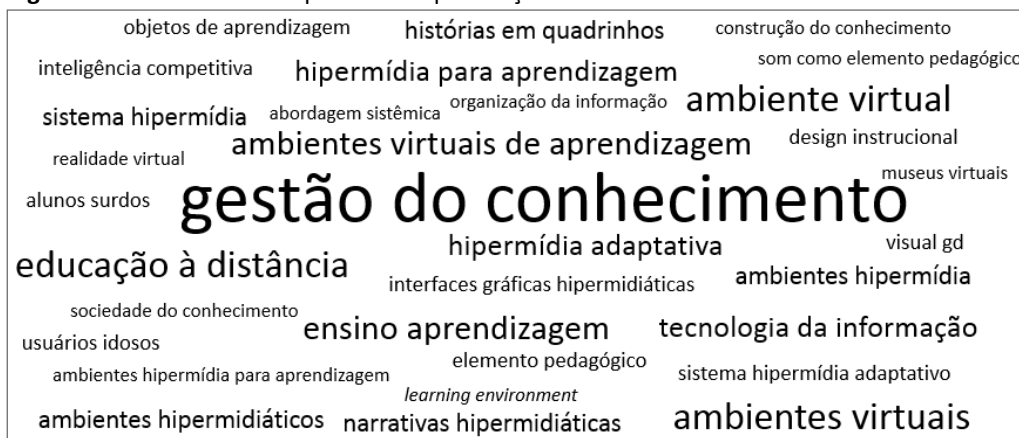
O termo mais frequente desta comunidade também é Gestão do Conhecimento (53). Em seguida vêm os termos Engenharia do Conhecimento (15) e Tecnologia da Informação (14). Os demais termos aparecem com baixa frequência.

Ao todo, os termos desta comunidade indicam uma atuação mais proeminente dos tópicos de estudo da área Engenharia do Conhecimento. Isso é compatível com a área de concentração dos três autores (*hubs*) que mais colaboram e, também, com a quantidade de autores docentes da área Engenharia do Conhecimento.

- Comunidade 4

Na Comunidade 4 foram encontrados 216 autores que representam 14,5% de toda a rede, com 498 produções. Os seis autores com maior centralidade de grau são V. R. U. (481), T. V. (338), J. A. S. (203), J. A. D. (201), A. T. C. P. (161) e A. F. B. (144). Alguns foram e outros são docentes do programa, sendo quatro autores da área Mídia e Conhecimento e dois autores da área Engenharia do Conhecimento. A Figura 6 exibe os termos mais frequentes encontrados nessa comunidade.

Figura 6 – Termos mais frequentes nas publicações da Comunidade 4



Fonte: elaborado pelos autores.

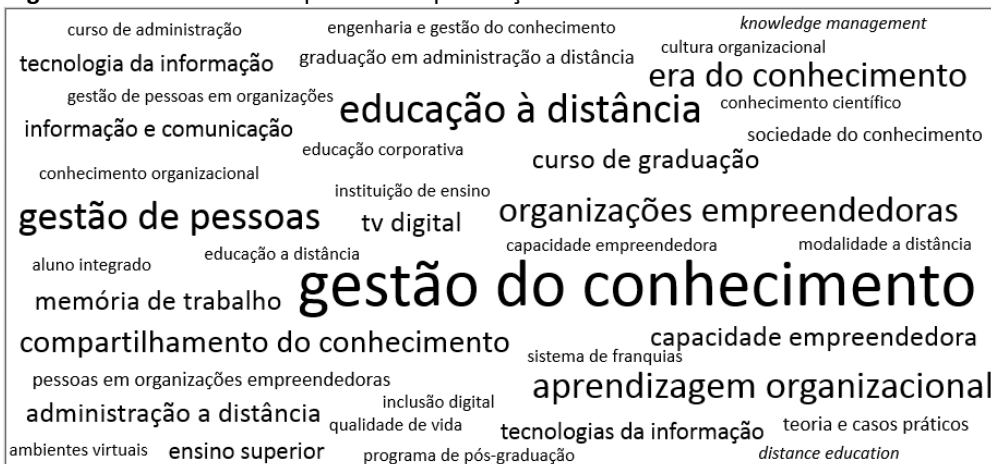
O termo *Gestão do Conhecimento* aparece como o termo mais recorrente desta comunidade (frequência igual a 20). Os termos *Ambientes Virtuais*, *Educação à Distância* e *Ambiente Virtual* aparecem logo em seguida com frequência igual a 17.

Observa-se que, ao todo, os termos desta comunidade representam a área de concentração da maioria dos *hubs*, *Mídia* e *Conhecimento*.

- Comunidade 5

Na Comunidade 5 foram encontrados 197 autores que representam 13,2% de toda a rede e publicaram 465 produções. O autor com maior centralidade de grau é M. K. N. (256), docente da área *Mídia e Conhecimento*. O segundo autor com maior centralidade é A. S. V. P. (227), discente do programa. A terceira, quarta e quinta posição no *ranking* de centralidade pertencem a F. J. S. (215) docente da área *Mídia e Conhecimento*, A. M. B. F. (210) e E. M. L. (174) ambos docentes da área *Gestão do Conhecimento*. A comunidade possui mais dois docentes da área de concentração *Mídia e Conhecimento*, totalizando seis pesquisadores (quatro da área *Mídia e Conhecimento* e dois da área *Gestão do Conhecimento*). A Figura 7 exhibe os termos mais frequentes encontrados nesta comunidade.

Figura 7 – Termos mais frequentes nas publicações da Comunidade 5



Fonte: Elaborado pelos autores

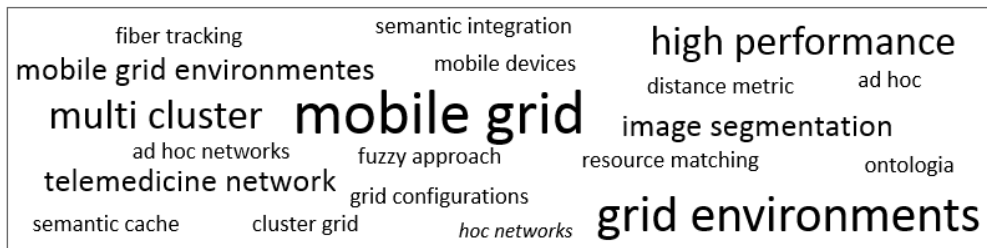
O termo mais frequente desta comunidade é Gestão do Conhecimento (45). Em segundo lugar vem o termo Educação à Distância (23), normalmente pesquisado na área Mídia e Conhecimento, e em terceiro lugar está o termo Gestão de Pessoas (12), seguido dos termos aprendizagem organizacional (11), organizações empreendedoras (10).

Nesse caso, verifica-se que não há uma convergência dos termos com a área de concentração da maioria dos *hubs* da comunidade, os quais são da área Mídia e Conhecimento.

- Comunidade 6

Na Comunidade 6 foram encontrados 115 autores que representam 7,7% de toda a rede, com 135 produções. Os três autores com maior centralidade de grau são A. v. W. (209), M. A. R. D. (173) e E. C. (76). O primeiro autor é ex-docente do programa e pertence a área de concentração Engenharia do Conhecimento. O segundo é docente do programa e pertence à mesma área de concentração. O último autor é participante externo ao programa. A Figura 8 exibe os termos mais frequentes encontrados nesta comunidade.

Figura 8 – Termos mais frequentes nas publicações da Comunidade 6



Fonte: Elaborado pelos autores

O termo mais frequente desta comunidade é *Mobile Grid* (7). Em seguida vem os termos *Grid Environments* (6) e *Multi Cluster* e *High Performance* com frequência igual a 5. Esses termos são representativos das linhas de pesquisa dos *hubs* desta comunidade.

Os termos desta comunidade destacam-se por estarem na língua inglesa. Um indicador dessa característica é que os principais *hubs* possuem um vínculo internacional estreito, com predominância de produções intelectuais advindas de publicações em periódicos e eventos internacionais.

- Comunidade 7, 8 e 9

Foram encontradas 3 comunidades isoladas, as quais juntas, representam 1% do total da rede. A Comunidade 7, com 11 autores, representa 0,7% da rede e tem o docente L. O. P. da área Gestão do Conhecimento como autor com maior centralidade de grau (14). Esta comunidade possui 22 produções. A tabela 2 exibe os termos mais frequentes encontrados na Comunidade 7.

Tabela 2 – Termos mais frequentes nas publicações da Comunidade 7

Frequência	Termos mais frequentes
11	[propriedade intelectual]
3	[inovação no agronegócio]

Fonte: Dados da pesquisa, 2013

A Comunidade 8 possui três autores e representa 0,2% da rede. Ela tem o docente P. C. B. N. da área Gestão do Conhecimento com a mesma centralidade de grau (2) dos outros autores. Nesta comunidade foram desenvolvidas sete produções e não foram encontrados termos com frequência maior que um. A análise manual dos títulos dessas produções revelou que a comunidade pesquisou tópicos como sustentabilidade e educação ambiental, os quais são mais relacionados à área Gestão do Conhecimento.

A Comunidade 9 é composta somente pelo autor N. L. L. que foi docente do programa e desenvolveu cinco produções no seu período de atuação. Nela também não foram encontrados termos com frequência maior que um. Porém a análise manual dos títulos das produções revelou publicações mais relacionadas à área Mídia e Conhecimento.

Como fechamento da análise das comunidades de colaboração identificadas neste estudo, a Tabela 3 exibe as comunidades de colaboração com as áreas de concentração predominantes, onde os termos das publicações são aderentes às áreas de concentração dos *hubs* das comunidades.

Tabela 3 – Classificação de comunidades de colaboração pelas áreas de concentração predominante

N. da comunidade	Nodos/autores	N. de publicações	Área de concentração predominante
1	358	564	Gestão do Conhecimento
2	324	652	Mídia e Engenharia do Conhecimento
3	263	436	Engenharia do Conhecimento
4	216	498	Mídia e Conhecimento
5	197	465	-
6	115	135	Engenharia do Conhecimento
7	11	22	Gestão do Conhecimento
8	3	7	Gestão do Conhecimento
9	1	5	Mídia e Conhecimento

Fonte: Dados da pesquisa, 2013

Como se pode observar na Tabela 3, a única comunidade onde os termos das publicações não são aderentes as área de concentração dos *hubs* é a Comunidade 5. Nessa comunidade a maioria dos termos é da área Gestão do Conhecimento, porém, a área da maioria dos *hubs* é Mídia e Conhecimento. Uma possível explicação para essa situação é que os autores da área estão pesquisando tópicos ou usando termos que são mais aderentes a área Gestão do Conhecimento.

Além disso, a Tabela 3 indica que a Comunidade 1 possui o maior número de nodos e tem 564 publicações. A Comunidade 2, com 324 nodos, ou seja, 34 nodos a menos que a Comunidade 1, possui, contudo, um número maior de publicações (652), 88 publicações a mais que a Comunidade 1. O mesmo ocorre com as comunidades adjacentes 3 e 4 (pelo número de nodos em ordem decrescente), em que a Comunidade 3 tem 263 nodos e 436 publicações e a Comunidade 4 tem 216 nodos, mas com 498 publicações, isto é, 62 publicações a mais que a Comunidade 3. Essas informações evidenciam que não há uma associação positiva entre o número de autores que colaboram em uma comunidade com o número de publicações.

5 DISCUSSÃO

O objetivo deste trabalho foi identificar e analisar as comunidades de colaboração científica do PPGECC. Tal motivação se deve a: (i) colaboração científica como um fator importante para aumentar o impacto das produções intelectuais – requisito este com significativo grau de impacto nas avaliações dos programas de pós-graduação *stricto sensu* no

Brasil (WEISZ; ROCO, 1996; BALANCIERI et al., 2005; ROSAS, 2013); (ii) a colaboração científica como importante mecanismo para tornar os diálogos mais amplos, uma vez que as produções intelectuais dos programas de pós-graduação interdisciplinar devem visar um saber cujos limites ultrapassam as fronteiras disciplinares (CAPES, 2013).

O PPGEHC foi um dos primeiros programas de pós-graduação, *stricto sensu*, interdisciplinar no Brasil. Seu conceito tem evoluído ao longo das avaliações da CAPES, e isso se deve, principalmente, ao número de produções intelectuais produzidas pela colaboração docente-discente. Essa parceria, além de ter uma pontuação mais significativa perante a CAPES, proporciona maior qualidade nas produções devido à bagagem de conhecimentos e experiências que os docentes possuem ao longo de sua carreira acadêmica.

Analisando os resultados encontrados na seção anterior, percebeu-se que as comunidades de colaboração do PPGEHC detectadas se arranjam de formas distintas. Dessas comunidades, três delas estão isoladas, ou seja, não possuem vínculos de coautoria com autores das demais comunidades, entretanto, elas representam um percentual pouco significativo na rede (1%).

Alguns grupos colaboram com mais intensidade, como a Comunidade 4, com 216 autores, dentre os quais possuem as maiores centralidades de grau em relação aos autores das outras comunidades.

Cada comunidade fornece subsídios através de um corpo de termos que permeiam as três áreas de concentração do PPGEHC, para avaliar como os esforços de pesquisa estão divididos entre as áreas de concentração do programa. Além disso, fica evidente o papel dos *hubs* na concentração dos termos em uma área específica. Na Comunidade 1, em que a área de concentração predominante dos *hubs* é Gestão do Conhecimento, os termos estão mais aderentes a esta área. Na Comunidade 2 em que a área de concentração predominante dos *hubs* é Engenharia do Conhecimento e Mídia e Conhecimento, percebe-se um conjunto de termos aderentes a essas duas áreas com interações com a área Gestão do Conhecimento. Essa característica das comunidades vai ao encontro do que a CAPES (2013) determina e orienta aos programas de pós-graduação interdisciplinares, que é necessário romper as barreiras e os limites da disciplinaridade (isto fica evidente nas comunidades por meio dos diversos temas que se apresentam) e que as produções devem estar alinhadas e serem coerentes com as áreas de concentração dos docentes.

Uma característica que se destacou nas cinco primeiras comunidades é a aparição do termo Gestão do Conhecimento como o termo mais frequente da comunidade. Um possível indicador desta ocorrência é que as áreas Engenharia do Conhecimento e Mídia e Conhecimento encontram sinergia com a área Gestão do Conhecimento, por meio dos seguintes aspectos: pesquisa e desenvolvimento de metodologias de identificação, representação e gestão de conhecimento; aplicação de sistemas de conhecimento à gestão do conhecimento organizacional (formalização, memória e tomada de decisão); pesquisa da construção, comunicação e difusão do conhecimento por meio de sistemas interativos multimídia e suas relações com a gestão do conhecimento (EGC, 2014).

6 CONSIDERAÇÕES FINAIS

Este artigo analisou a rede de colaboração ou coautoria de produções bibliográficas do PPGEHC no período de 2005 a 2012, com o objetivo de identificar as comunidades de colaboração e suas áreas de concentração de pesquisa. Para isso foi utilizado o método de detecção de comunidades implementado no *software* de análise de rede social Gephi e uma técnica de extração de termos a partir do título das produções bibliográficas.

Por meio da análise dos resultados foi possível verificar a adequabilidade das pesquisas do PPGEHC às áreas de concentração dos *hubs*, que são ou foram docentes do programa. Esta

constatação é um fator de impacto importante para o PPGE GC uma vez que a produção intelectual é um dos requisitos de avaliação dos programas de pós-graduação brasileiro.

Além disso, os resultados obtidos nesse artigo são úteis para a tomada de decisão em nível operacional e gerencial, tanto por parte da comunidade acadêmica do PPGE GC quanto por parte da comunidade extra-acadêmica.

Em nível operacional, eles podem servir de guia para os novos integrantes da comunidade acadêmica do PPGE GC, tanto docentes quando discentes, encontrarem os grupos existentes e suas respectivas áreas de concentração, minimizando assim os esforços comumente demandados para encontrar tais grupos. O mesmo se aplica para a comunidade externa ao programa porque permite que sejam detectadas novas oportunidades de parcerias acadêmica e extra-acadêmica, processo de suma relevância para programas de pós-graduação interdisciplinar. Decisões tais como orientar os docentes representantes das comunidades isoladas a publicarem conjuntamente com outros autores de outras comunidades podem ser tomadas pela coordenação do programa.

Em nível gerencial eles podem servir de insumo para a coordenação do programa tomar decisões estratégicas, como por exemplo, fomentar o trabalho em conjunto dos grupos de Engenharia do Conhecimento e Mídia e Conhecimento de forma a criar uma comunidade com foco de pesquisa na interação dessas duas áreas.

Por fim, acredita-se que a aplicação do método de detecção de comunidades e extração dos termos mais frequentes dos títulos de produção utilizados nesse artigo possam ser úteis para avaliação das redes de colaboração de outros programas de pós-graduação.

A limitação que este artigo apresenta, e que se suprido pode gerar análises mais refinadas para um trabalho futuro, é que foi considerada apenas a extração de termos pelos títulos dos trabalhos. A extração de termos considerando o resumo e as palavras-chave possibilitam inferências mais concretas com relação às características das publicações das comunidades encontradas, bem como possibilita a elucidação do nível de cooperação entre as áreas de concentração.

Além disso, uma busca no portal de periódicos da CAPES com os termos colaboração científica ou redes de coautoria mostrou que há poucas pesquisas que analisaram as redes de colaboração científica em programas de pós-graduação e nenhuma dessas pesquisas teve como estudo de caso um programa interdisciplinar. Programas interdisciplinares têm características significativamente distintas de programas disciplinares. Assim, para este artigo, a discussão se limitou à análise dos resultados do caso do PPGE GC, deixando para pesquisas futuras uma análise comparativa entre programas interdisciplinares e disciplinares.

REFERÊNCIAS

ABBASI, A.; ALTMANN, J.; HWANG, J. Evaluating scholars based on their academic collaboration activities: Two indices, the RC-index and the CC-index, for quantifying collaboration activities of researchers and scientific communities. *Scientometrics*, v. 83, n. 1, p. 1-13, 2010.

BALANCIERI, R. et al. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. *Ciência da Informação*, v. 34, n. 1, 2005.

BARABÁSI, A. L. **Linked**: a nova ciência dos networks. São Paulo: Leopardo, 2009.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. In: **International AAAI Conference on Weblogs and Social Media**, 2009.

BLONDEL, V. D. et al. Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and Experiment**, v. 2008, n. 10, p. P10008, 2008.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Documento de área e Comissão da Trienal 2013. 2013. Disponível em: <http://www.capes.gov.br/component/content/article/44-avaliacao/4674-interdisciplinar>.

Acesso em: 20 maio 2015.

EGC. Engenharia e Gestão do Conhecimento. Áreas de Concentração. Disponível em: <http://www.egc.ufsc.br/index.php/pt/egc/pos-graduacao/programa/areas-de-concentracao>.

Acesso em: 20 maio 2015.

FELDMAN, R.; SANGER, J. **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data**. New York: Cambridge University Press, 2007. 424 p.

FERRAZ, Renato Ribeiro Nogueira; QUONIAM, Luc; ALVARES, Lillian Maria Araújo de Rezende. Avaliação de redes multidisciplinares com a ferramenta scriptlattes: os casos da nanotecnologia, da dengue e de um programa de pós-graduação Stricto Sensu em Administração. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 19, n. 40, p. 67-98, 2014.

FREEMAN, L. Centrality in Social Networks: Conceptual Clarification. **Social Networks**, v. 1, n. 3, p. 215–239, 1979.

GIRVAN, M.; NEWMAN, M. E. J. Community structure in social and biological networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 12, p. 7821-7826, 2002.

GLÄNZEL, W.; SCHUBERT, A. Analysing scientific networks through co-authorship. In: MOED, H. F. W.; GLÄNZEL, U.; SCHMOCH, A. (Eds.). **Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems**. Dordrecht: Kluwer Academic Publishers, 2004. p. 257-276.

HARA, N. et al. An emerging view of scientific collaboration: Scientists' perspectives on factors that impact collaboration. **Journal of the American Society of Information Science and Technology**, v. 54, n. 10, p. 952-965, 2003.

HENDRICKSON, B.; KOLDA, T. Graph partitioning models for parallel computing. **Parallel Computing**, v. 26, n. 12, p. 1519-1534, 2000.

KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research Policy**, v. 26, n. 1, p. 1-18, 1997.

KOK, Daniel de; BROUWER, Harm. Natural Language Processing for the Working Programmer. 2011. Disponível em: <http://nlpwp.org/book/>. Acesso em: 20 maio 2015.

MARTINS, D. L.; FERREIRA, S. M. S. P. Proposta de metodologia de mapeamento e avaliação da produção científica da Universidade de São Paulo com foco na estrutura e dinâmica de suas redes de colaboração científica: em busca de novos modelos causais. **Liinc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 181-195, 2013.

MARTINS, Dalton Lopes. **Análise de redes sociais de colaboração científica no ambiente de uma federação de bibliotecas digitais**. 2012. Tese (Doutorado) - Curso de Programa de Pós-graduação em Ciência da Informação, Universidade de São Paulo, São Paulo, 2012.

- MILOJEVIC, S. Modes of Collaboration in Modern Science: Beyond Power Laws and Preferential Attachment. **Journal of the American Society for Information Science and Technology**, v. 61, n. 7, p. 1410-1423, 2010.
- NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 2, p. 404-409, 2001.
- _____. Modularity and community structure in networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 23, p. 8577-8582, 2006.
- NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical Review E**, v. 69, n. 026113, p. 1-16, 2004.
- PACHECO, R. C. S.; TOSTA, K. C. B. T.; FREIRE, P. S. Interdisciplinaridade vista como um processo complexo de construção de conhecimento: uma análise do Programa de Pós-Graduação EGC/UFSC. **RBPG: Revista Brasileira de Pós-Graduação**, v. 7, n. 12, p. 136-159, 2010.
- PINEAU, G. Recherches transdisciplinaires et Université. In: PAUL, P; PINEAU, G. (Org.). **Transdisciplinarité et formation**. Paris: L'Harmattan, 2005. p. 11-27.
- POMBO, O. Epistemologia da Interdisciplinaridade. **Revista Ideação**, v. 10, n. 1, p. 09-40, 2008.
- ROSAS, Fábio Sampaio. **Indicadores de impacto, visibilidade e colaboração para a produção científica da Pós-graduação brasileira: um estudo nos programas de excelência na área de Zootecnia**. 2013. Dissertação (Mestrado) - Curso de Mestrado em Ciência da Informação, Universidade Estadual Paulista, Marília, 2013.
- SCOTT, J. **Social Network Analysis: A Handbook**. 2. ed. London: SAGE Publications, 2000.
- VANZ, S. A. S.; STUMPF, I. R. C. Colaboração científica: revisão teórico-conceitual. **Perspectivas em Ciência da Informação**, v. 15, n. 2, p. 42-55, 2010.
- WASSERMAN, S.; FAUST, K. **Social Network Analysis: methods and applications**. Cambridge: Cambridge University Press, 1994. 857 p.
- WEISZ, J.; ROCO, M. C. **Redes de pesquisa e educação em engenharia nas américas**. Rio de Janeiro: FINEP, 1996.

Artigo recebido em 22/12/2014 e aceito para publicação em 07/06/2015
