

Rede de Pesquisadores Brasileiros em Ontologia: Uma Análise de Rede Social

Andréa S. Bordin¹, Alexandre Leopoldo Gonçalves¹

¹Departamento de Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brazil

andreabord@gmail.com, a.l.goncalves@ufsc.br

Abstract. *The research interest in the topic ontology has grown considerably in recent decades. Ontologies are searched or used primarily in the areas of Humanities, Exact and Health. This article aims to analyze the collaboration network based on co-authorship of Brazilian researchers on this topic. We retrieved 1179 articles in journals and national and international conferences from the Scopus database. The co-authorship network was created and analyzed using metrics social network analysis. The results show the existence of a fragmented network with many individual components, where the major component of the area presents authors with higher computing centralities.*

Resumo. *O interesse de pesquisa pelo tópico ontologia vem crescendo consideravelmente nas últimas décadas. Ontologias são pesquisadas ou utilizadas principalmente nas áreas de Ciências Humanas, Exatas e Saúde. Esse artigo objetiva analisar a rede de colaboração baseada em coautoria dos pesquisadores brasileiros nesse tópico. Foram recuperados 1179 artigos em periódicos e conferências nacionais e internacionais a partir da base de dados Scopus. A rede de coautoria foi criada e analisada através de métricas de análise de rede social. Os resultados mostram a existência de uma rede fragmentada com muitos componentes isolados, onde o maior componente apresenta autores da área de computação com as maiores centralidades.*

1. Introdução

Ontologia vem se tornando um tópico de interesse de pesquisadores em todo o mundo. Inicialmente abordado por pesquisadores da área de Ciências Humanas pelas suas raízes na Filosofia, posteriormente se transformou em objeto de pesquisa da área de Computação.

Uma pesquisa recente pelo termo “ontolog*” na base de dados Scopus revelou a existência em torno de 56.000 documentos contra 2 documentos em 1980. A mesma pesquisa, para documentos onde pelo menos um autor é brasileiro encontrou 1179 documentos, sendo o primeiro deles publicado no ano de 1995.

A partir desse ano, o número de artigos com a participação de pelo menos um autor brasileiro teve uma ascendência constante e atingiu seu pico em 2010, com 198 documentos, o que denota o interesse e a importância desse tópico de pesquisa para a ciência no Brasil.

Esse artigo analisa a colaboração científica entre os pesquisadores brasileiros em ontologia. Estudar a colaboração científica de um campo de conhecimento permite que sejam descobertos os focos e fluxos de transferência de conhecimentos e habilidades do grupo que colabora (KATZ; MARTIN, 1997). O objetivo deste trabalho é descobrir os autores que mais colaboram e as sub-redes de colaboração. O indicador utilizado é o de coautoria dos documentos coletados na base Scopus. Com isso, criou-se a rede de coautoria dos pesquisadores e o método de análise de rede social foi aplicado. Os resultados obtidos buscam fornecer o perfil da área de pesquisa em ontologia no Brasil e com isso ajudam novos e atuais pesquisadores a se posicionar nesse panorama.

A seguir será apresentado o referencial da área de análise de rede social, os procedimentos metodológicos, os resultados, a discussão e a conclusão do trabalho.

2. Análise de Rede Social

A modelagem de sistemas em rede vem sendo aplicada em áreas diversas como epidemiologia (Moore e Newman, 2000) e colaboração científica (Newman, 2004). 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. Uma rede pode ser direcionada ou não direcionada e as arestas podem ser valoradas ou não valoradas.

Segundo Katz e Martin (1997) a colaboração científica pode ser estudada segundo outros indicadores, porém a coautoria é o indicador mais utilizado. Logo, ao modelar uma rede de colaboração científica os vértices representam os autores e as arestas representam os artigos produzidos em parceria com outros autores. Esse tipo de rede é não direcionada e valorada porque a relação de coautoria é assíncrona e a valoração ocorre em função do número de artigos publicados em conjunto.

Dados modelados em rede são passíveis de serem analisados através de métricas de análise de rede social, a qual por sua vez, tem suas raízes na teoria de grafos. Segundo WASSERMAN e FAUST (1994) a área de análise de rede social (*social network analysis - sna*) tem atraído muito interesse nas últimas décadas. Através das métricas de *sna* é possível identificar aspectos, tais como: a) padrões de relacionamento entre os atores de uma rede; b) a conectividade entre os mesmos; c) a formação de *clusters*; d) a evolução da rede ao longo do tempo e, e) o fluxo de comunicação, informação e conhecimento dentro da rede.

Uma rede pode ser analisada segundo o escopo de estrutura, onde a medida de densidade é utilizada. Segundo Scott (2000), a densidade é um dos conceitos mais utilizados em teoria dos grafos, pois esta medida descreve o nível geral de ligações entre os pontos de um grafo. Um grafo "completo" é aquele em que todos os pontos são adjacentes um ao outro, ou seja, cada ponto é ligado diretamente a todos os outros pontos. Quanto mais pontos estão ligados uns aos outros, mais denso será o gráfico. No contexto de uma rede de coautoria a densidade reflete o percentual do total da rede com o qual um ator foi coautor de um artigo. (FISCHBACH; PUTZKE e SCHODER, 2011).

No escopo individual, existem algumas métricas de centralidade que procuram descrever as propriedades de localização de um ator na rede. Os atores mais importantes ou mais proeminentes estão normalmente localizados em posições estratégicas dentro da

rede (WASSERMAN; FAUST, 1994). A centralidade de um ator pode ser local ou global. A centralidade local está preocupada com a importância de um ator na sua vizinhança, enquanto que a centralidade global diz respeito a proeminência do ator dentro de toda a rede. A centralidade local é medida através da centralidade de grau (*degree centrality*), enquanto que a centralidade global é medida através da centralidade de intermediação (*betweenness centrality*) e centralidade de proximidade (*closeness centrality*).

A centralidade de grau de um ator corresponde ao número de arestas incidentes ou ao número de vértices adjacentes a ele. Segundo Freeman (1979) a centralidade de grau reflete a posição e o papel do ator em termos de popularidade e atividade. Em redes valoradas, onde a aresta possui um peso, a centralidade de grau pode levar em conta o valor ou peso da aresta. Em redes de coautoria essa medida determina o grau de colaboração de um ator.

A centralidade de proximidade é uma medida que indica a proximidade de um determinado ator em relação aos demais atores da rede, sendo definida pela soma das distâncias geodésicas entre um determinado vértice e todos os outros vértices do grafo (FREEMAN, 1979). Numa rede de coautoria, um autor com uma centralidade de proximidade alta pode indicar uma maior possibilidade de estabelecer parcerias de publicação na rede por estar mais próximo em relação a todos os outros autores (SOUZA; BARBASTEFANO; LIMA, 2012).

Por sua vez, a centralidade de intermediação mede o quanto um determinado ator se encontra "entre" os vários outros atores no grafo, ou seja, atribui importância a um ator em função do fluxo que passa por ele para interligar outros dois atores da rede através do menor caminho possível. Numa rede de coautoria, um autor com alto valor de centralidade de intermediação indica que um número significativo das parcerias estabelecidas na rede envolve, de forma direta ou indireta, as publicações relacionadas a esse ator (SOUZA; BARBASTEFANO; LIMA, 2012).

Uma das maiores preocupações de analistas de redes sociais é a identificação de subgrupos de atores dentro de uma rede. Subgrupos são subconjuntos de atores entre os quais existem laços fortes, diretos, intensos, frequentes ou positivos (WASSERMAN; FAUST, 1994). A identificação de componentes ou subgrafos dentro de uma rede é uma das técnicas para analisar uma rede do ponto de vista de um grupo de atores. Componentes são subgrafos que estão conectados dentro do grafo, mas desconectados entre os subgrafos. Se um grafo contém um ou mais pontos "isolados", esses pontos também são chamados de componentes. Componente gigante é o nome dado ao subgrafo que contém o maior número de atores conectados. Numa rede de coautoria a presença de mais de um componente na rede indica a existência de grupos que publicam isoladamente

3. Procedimentos metodológicos

Os procedimentos dessa pesquisa estão divididos em três etapas:

1) Coleta de dados: A base de dados utilizada na pesquisa foi a Scopus. Ela é considerada a maior base de dados de resumos, citações e textos completos da literatura científica mundial revisada, com cobertura desde 1960, com mais de 20.500 títulos de

aproximadamente 5.000 editoras internacionais e atualizações diárias (SCOPUS, 2013). Para a pesquisa foram recuperados todos os documentos com o termo “ontolog*” produzidos por ao menos um autor afiliado a uma instituição brasileira. Para isso foi utilizado o termo “ontolog*” nos campos *title*, *abstract*, *keywords* juntamente com o termo “brazil” no campo *affiliation country*. A pesquisa retornou 1179 documentos, os quais foram exportados para o formato .RIS.

2) Normalização dos dados: A normalização dos dados foi realizada através de um processo semiautomático, com a extração e ordenação dos nomes de todos os autores através de uma aplicação e a conferência manual das inconsistências nos nomes dos autores.

3) Análise de rede: a rede de coautoria foi criada por uma aplicação que analisou e contabilizou todas as coautorias dos documentos a partir dos documentos coletados e normalizados. A rede consiste de uma relação de nodos e uma relação de pares de nodos juntamente com o peso (número de artigos publicados em coautoria). A análise da rede foi efetuada com a utilização do software de análise exploratória de dados Gephi (BASTIAN; HEYMANN; JACOMY, 2009).

4. Resultados

Na rede de coautoria foram identificados 2738 autores e 12345 relações. A densidade encontrada foi 0,003 onde o valor máximo é 1.0 e o grau médio de colaboração foi 9.0. Os autores foram classificados segundo a medida de centralidade de grau que leva em consideração o peso das relações (C.G. c/ Peso) e que determina o grau de colaboração entre os atores da rede. A tabela 1 apresenta dez autores, sua posição no ranking e o grau de colaboração.

Tabela 1: Ranking dos pesquisadores brasileiros em ontologia

<i>Ranking</i>	Instituição	Autor	Nu. Doc.	C.G.	C.G. c/Peso
2	UFRJ	De Souza# J.M.	20	38	85
3	UFES	Guizzardi# G.	42	40	75
14	PUCRIO	Casanova# M.A.	14	30	56
61	UFPE	Freitas# F.	16	43	55
62	UFC	De Macedo# J.A.F.	8	28	54
119	UFAL	Bittencourt# I.I.	9	29	46
121	UFRJ	Xexeo# G.B.	11	23	45
122	UFAL	Costa# E.	9	26	44
135	UFC	Vidal# V.M.P.	7	17	40
139	UNICAMP	Medeiros# C.B.	17	22	38

A análise de rede revelou a existência de 348 componentes isolados, onde o maior componente (componente gigante) possui 941 autores e representa 34,37% da rede. A figura 01 apresenta o componente gigante com destaque para os autores com maior centralidade de intermediação, tais como Siqueira# S.W.M. (0,35), De Souza#

J.M. (0,32) e Breitman# K.K. (0,31). Em relação à centralidade de proximidade os autores com maiores graus são Siqueira# S.W.M., De Souza# J.M. e Carvalho, G., todos com 0.14.

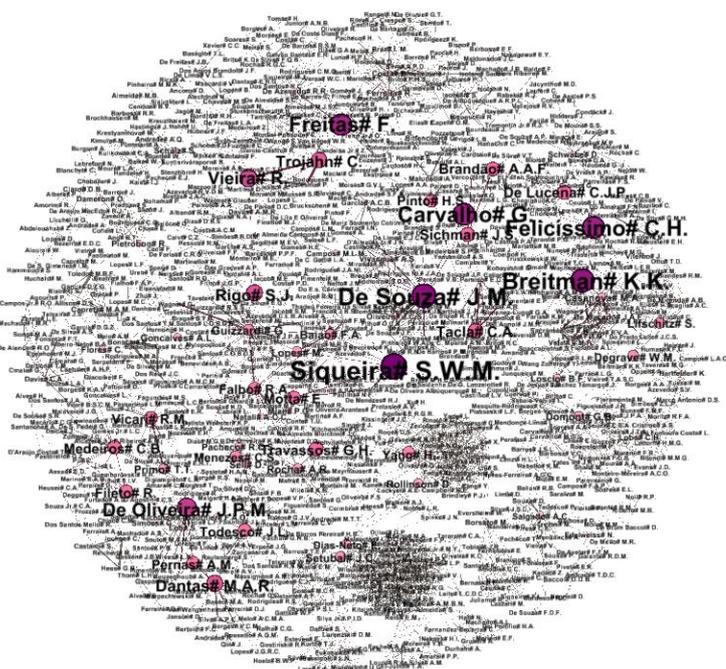


Figura 1: Componente gigante da rede de coautoria de pesquisadores em ontologia

O segundo maior componente encontrado na rede possui 110 autores e corresponde a 4% da rede. Esse componente foi formado principalmente pela publicação de cinco artigos que abordam o uso de ontologias na área de Bioinformática e foram coautorados por pesquisadores de diversos países. O terceiro e quarto maior componente, respectivamente com 59 e 54 autores, também são formados por autores com publicações nessa área. O quinto e o sexto componente contam com 51 e 44 autores respectivamente, cujas publicações são na área de ciência da computação.

5. Discussão e Conclusão

O ranking original dos autores com maior grau de colaboração apresentou muitos autores da área de Bioinformática. Verificou-se que os trabalhos desses autores abordam apenas o uso de ontologias conhecidas na área. Como os trabalhos não contribuem diretamente para o avanço da pesquisa em ontologia e o alto grau de colaboração encontrado nesses autores está mais relacionado com a coautoria de poucos documentos com muitos outros autores optou-se por não apresentá-los na tabela 1.

No ranking original de colaboração somente dois autores da área de Computação ocupam as primeiras dez posições. Isso pode evidenciar a necessidade de maior colaboração entre os pesquisadores dessa área.

A quantidade de componentes isolados (348) quando comparada ao número de autores da rede (2738) indica que a rede de coautoria dos pesquisadores brasileiros em

ontologia é bem fragmentada, ou seja, existem de grupos de pesquisadores trabalhando isoladamente ou sem colaboração entre grupos. A análise do segundo ao décimo componente isolado indica uma polarização das pesquisas entre as áreas de Computação e Bioinformática.

O maior componente de rede encontrado indica que 1/3 dos atores estão conectados por algum caminho. Nesse componente gigante, pesquisadores da área de Computação lideram o ranking de colaboração.

Esse trabalho apresentou uma análise preliminar da rede de pesquisadores brasileiros em ontologia, a partir da qual se pode concluir que existe espaço para uma maior colaboração entre os grupos de pesquisadores brasileiros e que as áreas de conhecimento mais representativas envolvidas com a pesquisa em ontologia são Computação e Bioinformática.

6. Referências

- BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media, 2009.
- FREEMAN, L. Centrality in Social Networks: Conceptual Clarification. *Social Networks*, 1(3), 215–239, 1979.
- FISCHBACH, K.; PUTZKE, J.; SCHODER, D. Co-authorship networks in electronic markets research. *Electronic Markets*, 21(1), 19–40, 2011.
- KATZ, J. S.; MARTIN, B. R. What is research collaboration? *Res. Policy*, 26(1):1-18, 1997.
- MOORE, C.; NEWMAN, M. E. J. Epidemics and percolation in small-world networks. *Phys. Rev. E* 61, 5678–5682, 2000.
- NEWMAN, M. E. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, v. 101, n. Suppl 1:5200-5205, april 2004.
- SCOPUS. Base de dados de literatura de pesquisa. Disponível em: <http://www.scopus.com>
- SCOTT, J. *Social Network Analysis. A Handbook*. 2nd edition. SAGE Publications: London, 2000.
- SOUZA, C. G.; BARBASTEFANO, R. G.; LIMA, L. S. Redes de colaboração científica na área de química no Brasil: um estudo baseado nas coautorias dos artigos da revista *Química Nova*. *Química Nova*, São Paulo, v. 35, n. 4, p. 671-676, 2012.
- WASSERMAN, S.; FAUST, K. *Social Network Analysis: methods and applications*. Cambridge University Press. Structural analysis in social the social sciences series, v. 8, (1994) 1999. 857 p. ISBN 0-521-38707-8.