

ESTRUTURAS DE HIERARQUIA E EVENTOS GERADORES EM REDES DE COLABORAÇÃO CIENTÍFICA:

estudo exploratório entre relações de coautoria e em bancas de defesas
de teses e dissertações

Eixo temático: Colaboração

Modalidade: Apresentação oral

1 INTRODUÇÃO

O estudo da estrutura e da dinâmica de formação das redes sociais tem fornecido importantes indícios para os estudos bibliométricos e cientométricos na Ciência da Informação (Marteleto, 2010). Os padrões de coautoria e de citação estão entre os trabalhos mais frequentes utilizando a análise de redes, até mesmo pela maior disponibilidade de bancos de dados que podem ser utilizados pelos pesquisadores. Duas propriedades frequentemente utilizadas para caracterizar o padrão estrutural de uma rede social é o fato de muitas delas formarem níveis de agrupamento dos nós acima do que seria esperado em uma rede aleatória, o que pode ser percebido pela análise do coeficiente de clusterização, e a distribuição de grau de centralidade possuir características de uma lei de potência e não uma distribuição normal, caracterizando uma rede de livre escala (Newman, 2010, pag. 263).

Pesquisas indicam que redes que possuem ao mesmo tempo alto grau de formação de grupos e distribuição de grau descrita por lei de potência denotam um tipo de organização hierárquica, ou seja, alguns nós mais centrais dessas redes concentram conexões em diversos grupos que são altamente conectados, centralizando o nível estrutural de conectividade em torno de si (Ravasz e Barabasi, 2003). Esse nível hierárquico pode ser quantificado usando a relação descoberta por Dorogovtsev, Goltsev e Mendes (2002), onde em redes de livre escala o coeficiente de clusterização (C) de um nó se relaciona com o seu número de conexões k da seguinte forma: $C(k) \approx k^{-1}$.

Neste artigo, iremos descrever e comparar a estrutura hierárquica de dois tipos de redes constituídas por pesquisadores da área da Ciências da Comunicação, sendo elas uma rede formada pelas relações de coautoria entre dois ou mais autores em periódicos científicos

e uma rede formada pelas relações de participação em bancas de defesa de teses e dissertações, seja de orientadores, orientandos e membros da banca. A partir da relação hierárquica encontrada em cada uma, descrevemos possíveis características dos eventos geradores de tais redes que possam explicar como são formadas e que efeitos causam em sua estrutura e dinâmica.

2 RESUMO DO CASO

O portal Univerciencia.org é uma biblioteca digital federada com foco específico na área das Ciências da Comunicação. Possui 49 fontes de informação de revistas científicas e 12 bibliotecas digitais de teses e dissertações de programas de pós-graduação específicos da área. No período de nossa extração de dados, realizada em dezembro de 2011, o banco de dados registrava 11.825 documentos disponíveis para consulta, sendo distribuídos entre 9.864 (83,4%) artigos de revistas científicas e 1.961 (16,6%) teses e dissertações.

Em termos de abrangência de pessoas que publicaram algum conteúdo que tenha sido registrado pela biblioteca, temos, após um processo de normalização, 9.587 nomes distintos de pesquisadores na base de revistas científicas e 2.465 nomes distintos de participantes registrados nas bibliotecas digitais de teses e dissertações. Além disso, os dados mais antigos de bancas de defesa são de 25 anos atrás e de 15 anos no caso das revistas científicas. Pelos números expostos acima, a Univerciencia.org é considerado o maior repositório brasileiro na área de Ciência da Comunicação.

É importante ressaltarmos que nem todas as bibliotecas digitais coletadas disponibilizam dados dos membros participantes da banca, sendo que apenas 4 das 12 bibliotecas continham esses dados e o restante das 8 bibliotecas disponibilizavam apenas dados do orientador e orientando.

3 METODOLOGIA

A pesquisa realizada é de cunho quantitativo tendo sido utilizado o banco de dados da biblioteca digital federada Univerciencia.org como amostra de dados para as análises realizadas, sendo estas organizadas em dois tipos: uma descrição da diferença em termos de estrutura hierárquica nas redes de coautoria e de relações em bancas de defesa, além de uma descrição de eventos que ajudam a retratar como se deu a formação dessas redes e que são

possíveis fatores a serem levados em consideração na explicação das diferenças estruturais das redes em análise.

Os dados foram coletados diretamente do banco de dados do site Univerciencia.org. Coletamos apenas os metadados das revistas científicas e bibliotecas digitais de teses e dissertações. Os metadados são padronizados em suas fontes de origem no padrão Dublin Core Simplificado, o que nos permitiu analisarmos apenas os campos *dc:author*, no caso das revistas científicas, e também *dc:contributor*, no caso dos documentos de teses e dissertações. Esses campos indicavam, quando disponíveis, os nomes de todos os coautores de um artigo científico e do autor de uma tese ou dissertação, além de seu orientador e participantes da banca de defesa. Os nomes identificados foram normalizados.

De modo a analisarmos a presença ou não de estruturas hierárquicas nas redes de coautoria e participação em bancas, realizamos uma distribuição do coeficiente de clusterização pelo grau de centralidade. Para facilitar a identificação de uma possível correlação, conforme vimos anteriormente, entre esses dois indicadores de rede, transformamos o gráfico da distribuição em escala logarítmica e calculamos a regressão linear da distribuição. Complementando, analisamos a distribuição de grau de centralidade dos nós no componente principal da rede (a parte da rede que contém a maior quantidade de nós interligados) de modo a identificar na maior estrutura da rede como essa distribuição pode ter relação com a hierarquia dos nós.

Para entendermos melhor que possíveis eventos poderiam explicar a organização hierárquica e estrutural dessas redes, analisamos a quantidade de anos que um nó é presente na rede, além da estrutura de agrupamento que ocorre entre os nós quando eles entram na rede, ou seja, se entram conectados a apenas mais uma outra pessoa, a mais duas, a três e assim por diante.

4 RESULTADOS E DISCUSSÃO

A distribuição do coeficiente de clusterização pelo grau de centralidade é apresentado na figura 1, a seguir.

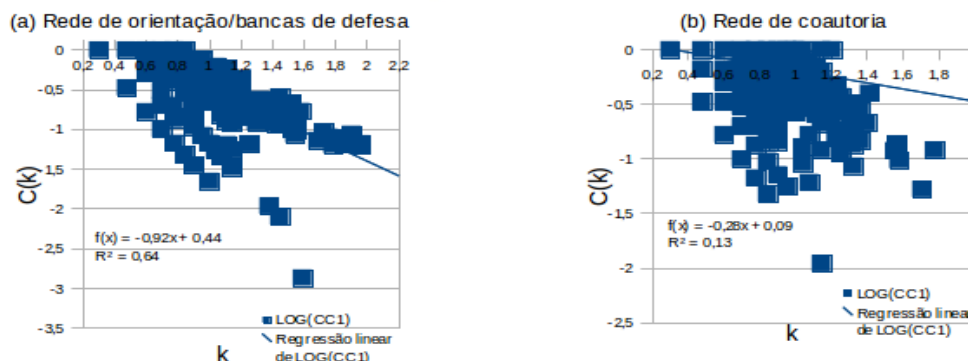
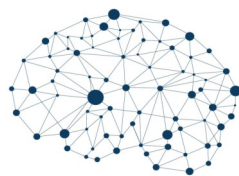


Figura 1. Distribuição logarítmica do coeficiente de clusterização por grau de centralidade. (a) Rede de bancas de defesa. (b) Rede de coautoria.

Observamos na figura 1a a distribuição da rede de bancas de defesa, que apresenta um coeficiente de regressão linear dos dados dispostos em escala logarítmica de 0,64, sendo considerada uma correlação positiva moderada. Já na figura 1b observamos a distribuição da rede de coautoria, apresentando um coeficiente de correlação de 0,13, sendo considerada uma correlação fraca. Desse modo, notamos indícios de uma estrutura hierárquica na rede de bancas de defesa, o que não é percebido na rede de coautoria em revistas científicas.

Analisando apenas a distribuição de grau dos nós pertencentes ao componente principal de cada uma das redes, percebemos importantes diferenças que denotam características estruturais dessas redes, conforme vemos na figura 2.

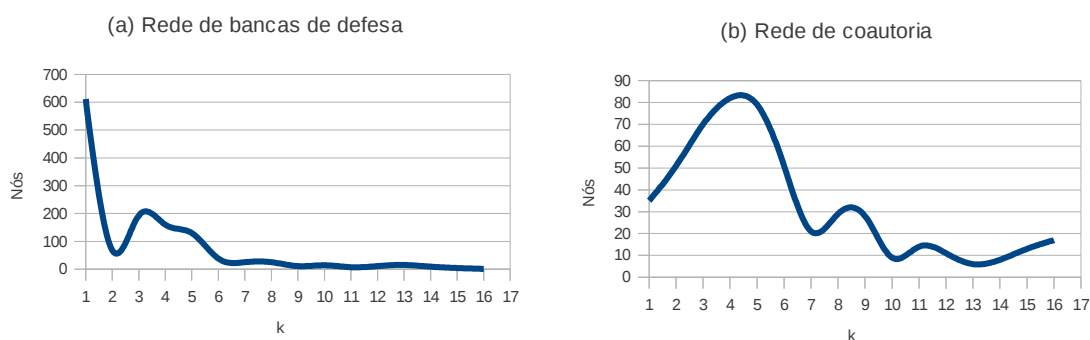


Figura 2. Distribuição de grau de centralidade do componente principal. (a) Rede de bancas de defesa. (b) Rede de coautoria.

A distribuição de grau de centralidade da rede bancas apresenta um maior número de nós com grau 1, indicando que esses nós estão conectados a apenas um outro nó no mesmo componente. Já a rede de coautoria apresenta um padrão diferente, sendo seu pico em torno do valor 4, indicando que há mais nós conectados a outros 4 do que outras configurações de conexões. Vale ressaltarmos que essa diferença entre a distribuição grau de centralidade dos nós indica uma maior concentração das conexões em menos nós na rede de bancas de defesa relativamente a rede de coautoria. Quando analisamos a quantidade de nós na rede coautoria

que possui até grau 5 encontramos um valor de 57%, passando esse valor para 83,2% na rede bancas de defesa, confirmando essa maior concentração de conectividade que é um dos fatores influentes na definição de hierarquia da própria rede.

Para avaliarmos possíveis eventos que levam a formação estrutural das redes em análise, apresentamos na figura 3 a distribuição relativa da quantidade de nós por anos em que estiveram presentes nas duas redes.

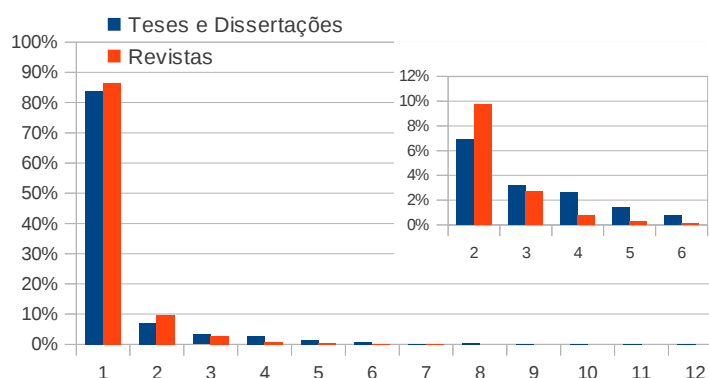


Figura 3. Distribuição de nós da rede por quantidade de anos que apareceram em alguma relação com outro nó. Observação: o gráfico menor na figura é uma ampliação, por fatores de escala, dos anos 2 a 6.

Observamos que até 02 anos, há relativamente mais nós da rede de coautoria presentes. Porém, quando analisamos de 03 anos em diante, percebemos que a rede de bancas de defesa apresenta uma maior quantidade relativa de nós atuantes. De certo modo, há nós com maior perenidade temporal na rede de bancas, chegando mesmo a encontrarmos nós com 12 anos de presença, sendo que o máximo encontrado na rede de coautoria foi de 06 anos. O fato de termos nós atuantes durante mais tempo numa rede é um dos fatores que devemos levar em consideração quanto a concentração de links que esses nós podem efetuar.

Além do tempo de presença, analisamos a seguir na figura 4 que tipos de associação com outros nós caracterizam a entrada de um novo elemento na rede. Por entrada, consideramos o primeiro evento que traz um nó para dentro da rede, sendo esse evento a publicação de artigo em coautoria com outros nós ou participação em uma banca de defesa. Vale ressaltar, em termos de amostra de tempo, que coletamos o mesmo período para as duas redes, o que equivale aos anos 10 a 25 da rede de bancas de defesa e aos anos 1 a 15 da rede de coautoria.

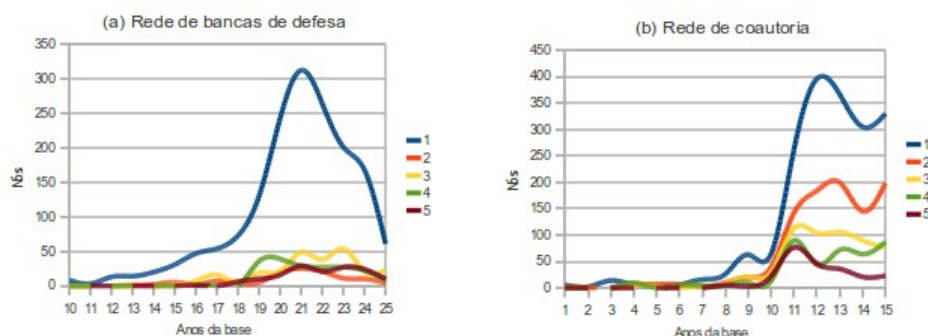
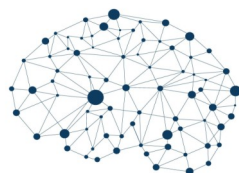


Figura 4. Distribuição dos modos de entrada dos nós na rede por ano da base. (a) Rede de bancas de defesa. (b) Rede de coautoria.

Observamos que a rede de bancas apresenta como característica uma entrada majoritária de nós associados a apenas um nó, sobretudo a partir do ano 18 até 24. Há um comportamento diferente em relação a rede de coautoria, onde partir do ano 10 há um aumento significativo de vários modos de associação, sendo que a associação a apenas um outro nó continua sendo o maior evento da rede, mas é seguido em crescimento por entradas de nós associados a mais outros 2, 3, 4 e 5. O fato de termos proporcionalmente mais nós associados a apenas um outro nó em seu evento de entrada na rede de bancas em relação a rede de coautoria é um fato que favorece novamente a concentração de links nos nós mais centrais da rede.

Na rede de coautoria, os eventos de entrada retratam pequenos grupos que iniciam juntos, o que percebemos menos na rede de bancas, sendo ali mais influente a relação em duplas, o que retrata a entrada de orientandos associados a seus orientadores na rede. É portanto, a partir das características retratadas neste artigo em relação a estrutura hierárquica das redes e aos eventos de entrada que propomos na figura 5 uma caracterização de dois diferentes tipos de eventos, responsáveis pela formação dos dois padrões de rede aqui analisados.

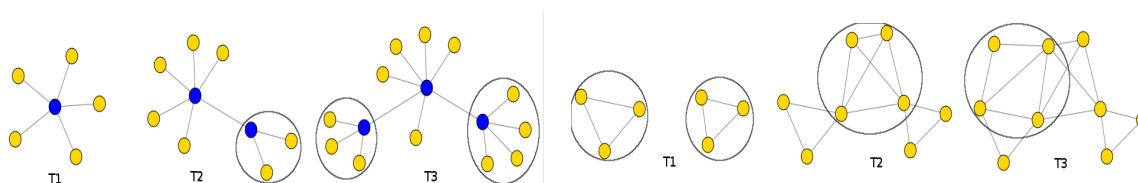


Figura 5. Eventos geradores. (a) Rede de bancas de defesa. (b) Rede de coautoria.

Retratamos três intervalos de tempo genéricos, T1, T2 e T3 para as duas redes. Na rede de bancas em azul temos o papel do orientador e em amarelo dos orientandos. Os eventos geradores da rede, em sua maioria, se dão quando na defesa de uma tese ou dissertação o nome do orientando aparece relacionado ao seu orientador. A partir do tempo T2, temos que

orientandos de um orientador passam a orientar trabalhos, ampliando a rede novamente ao seu entorno. Esse tipo de padrão de conectividade gera a hierarquização da rede que apresentamos ao longo deste trabalho. Já na rede de coautoria, as entradas ocorrem por associação entre pequenos grupos, que podem se interligar mais facilmente quando um autor de um grupo publica com outros autores conectados a outros grupos. Esse padrão de conectividade favorece uma descentralização nas conectividades da rede e maior distribuição de links entre os nós participantes.

5 CONCLUSÃO

A dinâmica de formação de uma rede se dá quando determinados eventos relacionais ocorrem, proporcionando conexões e levando a uma formação estrutural que caracteriza um padrão específico. Neste artigo, estudamos, a partir dos indícios de diferentes padrões hierárquicos, os possíveis eventos geradores de redes de coautoria e participação em bancas.

A dinâmica relacional que cria a rede de participação em bancas é a relação orientador-orientando. Já o movimento na rede de coautoria segue outra dinâmica. O evento gerador se dá a partir de outras relações que não apenas a relação orientador e orientando. O efeito dos grupos de pesquisa, projetos e laboratórios acaba atuando aqui, mudando a relação de hierarquia na estrutura da rede. Os eventos geradores, por serem oriundos de motivações relacionais diferentes, criam diferentes impactos na estrutura da rede.

REFERÊNCIAS

DOROGOTSEV, Sergey, GOLTSEV, Alexander, MENDES, Jose Fernando. Pseudofractal Scale-free Web. **Phys Rev E Stat Nonlin Soft Matter Phys.** v. 65. p. 066122, june, 2002.

MARTELETO, Regina. Redes sociais, mediação e apropriação da informação: situando campos, objetos e conceitos na pesquisa em Ciência da Informação. **Pesq. bras. ci. inf.**, Brasília, v.3, n.1,p.27-46, jan./dez. 2010 .

NEWMAN, Mark E. J. **Networks: an introduction**. New York. Oxford Press. 2010.

RAVASZ, Erzsébet, BARABASI, Albert-László. Hierarchical organization in complex networks. **Phys Rev E Stat Nonlin Soft Matter Phys.** v. 67, p. 026112,feb, 2003.