

Análise de redes de coautoria por meio de redes semânticas uniformes

Inácio de Sousa Fadigas¹, Marcelo do Vale Cunha^{2,3}, Marcos Grilo Rosa^{1,4},
Hernane Borges de Barros Pereira^{2,4,5}

¹Universidade Estadual de Feira de Santana, Feira de Santana, BA.

²Programa de Modelagem Computacional, SENAI Cimatec, Salvador, BA, Brasil

³Instituto Federal da Bahia, Salvador, BA

⁴Programa de Doutorado Multiinstitucional e Multidisciplinar em
Difusão do Conhecimento - Universidade Federal da Bahia (Sede),
Salvador, Bahia - Brasil

⁵Universidade do Estado da Bahia (UNEB), Salvador, BA, Brasil

{fadigas,grilo}@uefs.br, celocunha@yahoo.com.br, hbbpereira@gmail.com

Abstract. *This paper aims to analyze networks of coauthors who publish in the same journal by means of semantic networks based on titles of their articles. The use of semantic networks allows a better understanding of how authors in the same journal are integrated, since a coauthorship network can be highly fragmented. Two journals have been chosen for implementing the method that consists of obtaining, by means of uniform semantic networks, an associated coauthorship network. The results suggest how researchers deal with the choice of words of titles of papers in keeping with the journal's identity.*

Resumo. *Neste trabalho pretende-se analisar redes de coautores que publicam em um mesmo periódico por meio de redes semânticas baseadas em títulos de artigos. O uso de redes semânticas possibilita uma melhor compreensão de como autores de um mesmo periódico se integram visto que uma rede de coautoria pode ser altamente fragmentada. Escolheu-se dois periódicos para aplicação do método que consiste em obter, por meio de redes semânticas uniformes, uma rede de coautoria associada. Os resultados apontam evidências de como pesquisadores lidam com a escolha das palavras de títulos de artigos condizentes com a identidade do periódico.*

1. Introdução

A colaboração científica é um trabalho conjunto de pesquisadores para alcançar o objetivo comum de produzir novo conhecimento científico [Katz and Martin 1997]. Neste sentido, a colaboração científica é um processo social e de interação entre pesquisadores cujas relações podem ocorrer de diversas formas [Vanz and Stump 2010]. Por ser invariante e verificável, coautoria tem sido utilizada para avaliar colaboração científica e o uso da Teoria de Redes tem contribuído de forma relevante para mapear a colaboração entre pesquisadores pertencentes a uma mesma comunidade científica.

[Newman 2001] analisou colaboração científica por meio de redes de coautoria, onde os atores da rede são pesquisadores e as relações são estabelecidas entre pesquisadores que publicaram pelo menos um artigo juntos. A partir de quatro banco de dados, sete redes de coautoria foram construídas cuja maior componente de cada uma delas possui pelo menos 57% dos vértices da rede. Segundo [Newman 2001], redes de coautoria tendem a ter mais relações sociais entre atores genuinamente conhecidos do que em uma rede de atores de filmes, por exemplo. A rede de coautores do periódico *Data and Knowledge Engineering* (DKE) foi analisada por [Chen et al. 2008] que também investigou a explosividade de sintagmas nominais extraídos de títulos e resumos de artigos publicados no periódico científico supracitado. Uma rede contendo estes sintagmas nominais e palavras-chave também foi construída.

Recentemente, [Carvalho et al. 2012] apresentou um método de análise de discursos em fóruns de redes sociais de modo que é possível identificar as ideias centrais discutidas. Desta forma, temas de interesse coletivos podem ser sugeridos. O método proposto por [Carvalho et al. 2012] explora e seleciona de forma eficiente o conteúdo extraído dos fóruns, através de uma técnica de organização automática dos textos postados que envolve tarefas como identificação de uma rede de co-ocorrência de termos e agrupamento hierárquicos de textos.

Redes formadas por pesquisadores que publicaram em um mesmo periódico podem ser altamente fragmentadas e a maior componente pode não ultrapassar 20% dos vértices da rede. [Rosa et al. 2012] mostraram que a integração dos autores em um contexto (i.e. periódico & programa pós-graduação) é um fator relevante na colaboração em termos de coautoria. Autores pertencentes a um mesmo programa de pós-graduação tendem a colaborar (em termos de coautoria) mais do que autores que publicam em um mesmo periódico ou grupo de periódicos cujas linhas editoriais são bastante próximas. Considerando que o contexto aqui é o de periódico e a fragmentação não permite uma análise mais aprofundada do discurso dos seus autores, partiu-se para a análise da rede semântica baseada nos títulos de seus artigos, como meio de integração dos atores desta típica rede social.

Neste trabalho, pretende-se analisar redes de coautoria com autores que publicam em um mesmo periódico por meio de redes semânticas baseadas em títulos de artigos de periódicos científicos (*RST*). Para cada rede de coautoria é possível associar uma rede semântica formada por palavras com significado intrínseco pertencentes aos títulos publicados por seus coautores. Um conjunto de palavras pertencentes a um mesmo título forma uma clique e duas cliques estão conectadas entre si se existir pelo menos uma palavra comum a dois títulos.

Dois periódicos foram objeto de investigação desta pesquisa e para cada um deles, construiu-se uma rede de coautoria e uma *RST* a partir dos títulos publicados por seus coautores. Em seguida, aplicou-se a estratégia de retirada uniforme e aleatória de vértices (*RR*) em cada *RST* e se percebeu a existência de um ponto crítico quando a fração de vértices removidos é igual a 70%, gerando redes semânticas uniformes. Neste ponto crítico, as palavras que não foram removidas estão presentes em média, em pelo menos 36% dos títulos enquanto que a retirada de 10% dos vértices implica que as palavras que não foram removidas estão presentes em média, em pelo menos 45% dos títulos. A partir da identificação dos títulos que contém as palavras presentes nas redes semânticas

uniformes, pode-se montar redes de coautoria associadas. Desta forma, esta pesquisa aponta evidências sobre o modo como autores lidam com a escolha das palavras dos títulos dos seus artigos considerando o escopo do periódico.

Este trabalho está organizado da seguinte forma: a seção 2 trata de redes de coautoria (subseção 2.1) e de redes semânticas (subseção 2.2); a seção 3 trata da metodologia empregada para determinação de uma rede de coautoria associada a uma rede semântica uniforme e dos índices da Teoria de Redes utilizados nesta pesquisa; a seção 4 contém uma análise dos resultados obtidos; na seção 5 algumas considerações são tecidas.

2. Trabalhos correlatos

2.1. Redes de Coautoria

[Katz and Martin 1997] afirmam que a avaliação de colaboração científica por meio de coautoria é um método relativamente prático e barato além de ser invariante e verificável. Por outro lado, [Vanz and Stump 2010] consideram que o contato direto (mesmo virtual) entre duas ou mais pessoas é a unidade fundamental da colaboração. Neste sentido, redes de coautoria são redes sociais visto que é altamente provável que há um contato direto entre coautores.

Coautoria foi objeto de investigação de [Newman 2001] que utilizou a Teoria de Redes. Uma rede complexa ou social pode ser modelada por grafos cuja estrutura constituiu-se de um conjunto V de vértices e um conjunto ϵ de arestas. Neste trabalho, todos os vértices das redes investigadas são do mesmo tipo (redes de 1-modo), as arestas não possuem direção (redes não-dirigidas), cada aresta conecta exatamente dois vértices e para cada par de vértices existe no máximo uma aresta conectando ambos. No caso de redes de coautoria, os vértices são os autores e uma aresta entre autores é estabelecida se eles publicaram um artigo em coautoria.

[Newman 2001] construiu redes de coautoria em Física, pesquisa biomédica e Ciência da Computação, encontrando padrões de colaboração aparentemente diferentes. Um mapeamento da autoria, coautoria e dos temas discutidos no periódico *DKE* foi realizada por [Chen et al. 2008]. Para tanto, [Chen et al. 2008] utilizou diversos recursos dentre os quais destacam-se sobreposições da rede de coautoria em mapas geoespaciais, rede de citações, mapas conceituais e redes de sintagmas nominais e palavras-chave. A propriedade mundo pequeno foi encontrada na rede de coautores do periódico *DKE*.

Redes de coautoria foram construídas por [Quintella et al. 2009] para analisar a dinâmica de aquisição de conhecimento em três universidades públicas do estado da Bahia. Métodos da Teoria de Redes foram utilizados para avaliar a produção científica em química em três períodos: amadurecimento (1994 – 1998); consolidação (1999 – 2003); qualificação (2004 – 2008). Além de detectar a presença do fenômeno mundo pequeno, [Quintella et al. 2009] também apreciou a influência da rede de coautoria em química da Universidade Federal da Bahia em duas outras universidades públicas baianas.

A partir de dados obtidos da Plataforma Lattes, [Laender et al. 2011] introduziram uma arquitetura para construir redes de coautoria a partir de uma perspectiva individual. Algoritmos específicos foram utilizados para coletar e extrair os dados e também para a visualização das redes construídas. [Lopes et al. 2011] propuseram novos procedimentos para avaliar a qualidade de programas de pós-graduação por meio de redes de coautoria.

A análise mostrou que pesquisadores de programas bem avaliados tendem a apresentar um comportamento colaborativo.

[Rosa et al. 2012] investigou redes de coautoria de um programa de pós-graduação em modelagem computacional e de periódicos em Educação Matemática por meio da abordagem por cliques [Fadigas and Pereira 2013]. A Análise de Redes Sociais é comumente empregada para compreender as relações entre os atores da rede. Na abordagem por cliques, prioriza-se o exame das relações entre as cliques, ou seja, das relações entre os grupos de autores de um mesmo artigo científico. Novos índices de coesão foram propostos por [Rosa et al. 2012] para entender as redes de coautoria analisadas.

2.2. Redes Semânticas

Nesta pesquisa, uma rede semântica é um sistema de representação do conhecimento baseado em grafos, onde os vértices são palavras e as arestas são relações entre as palavras. As redes semânticas investigadas neste trabalho foram construídas a partir do método desenvolvido por [Pereira et al. 2011] que construíram redes semânticas baseadas em títulos de artigos de periódicos científicos. Os títulos passaram por um tratamento que consiste basicamente em eliminar palavras sem significados intrínsecos e, em seguida, alterar as palavras restantes para a sua forma canônica. Após o tratamento, todas as palavras em um mesmo título estão conectadas entre si, formando uma clique. Assim, formou-se para cada periódico uma rede semântica de cliques.

Redes semânticas de cliques foram analisadas por [Caldeira et al. 2006] e por [Teixeira et al. 2010]. A partir de 312 textos literários, [Caldeira et al. 2006] construíram redes semânticas baseadas em discurso escrito, enquanto [Teixeira et al. 2010] construíram redes a partir de 12 discursos orais. Em ambos os trabalhos, eliminou-se palavras sem significado intrínseco e as relações entre duas palavras foram estabelecidas caso as palavras pertencessem à mesma sentença. Logo, cada sentença gerou uma clique e de cada discurso criou-se uma rede de cliques.

[Cancho and Solé 2001], [Caldeira et al. 2006] e [Teixeira et al. 2010] investigaram redes semânticas que apresentaram a propriedade mundo pequeno [Watts and Strogatz 1998] e a característica livre de escala [Barabási and Albert 1999]. [Bales and Johnson 2006] mostraram que redes semânticas derivadas de linguagem natural tendem a ser redes mundo pequeno e livre de escala. As redes semânticas baseadas em títulos de periódicos científicos construídas por [Pereira et al. 2011] apontaram a presença do fenômeno mundo pequeno e a distribuição de graus evidenciaram um regime de lei de potência.

3. Metodologia

Neste experimento, duas redes de coautoria de dois periódicos científicos publicados em inglês foram investigadas: *Agricultural and Forest Entomology - AFE*; *Probabilistic Engineering Mechanics - PEM*. Os critérios de escolha para os periódicos foram: fator de impacto maior do que 1; disponibilidade dos artigos na Internet; cada periódico representar tão bem quanto possível uma área de conhecimento, incluindo áreas multidisciplinares. Os dados coletados dos periódicos *AFE* e *PEM* compreendem desde o primeiro volume publicado até, respectivamente, os anos de 2008 e 2009. Para cada periódico construiu-se uma rede de coautoria onde os vértices são autores e as arestas são estabelecidas se dois

autores publicaram um artigo juntos no periódico em questão. As redes de coautoria dos periódicos *AFE* e *PEM* possuem elevada fragmentação ou, de forma equivalente, baixa conectividade.

O conceito de conectividade (C_{on}) de uma rede com N vértices foi inicialmente proposto por [Krackhardt 1994]:

$$C_{on} = 1 - 2 * \frac{PNC}{N(N - 1)} \quad (1)$$

onde PNC significa o número de pares não conectados. De outro modo, a conectividade (C_{on}) é a razão entre o número de pares conectados e o número de pares possíveis da rede. [Borgatti 2006] usou o conceito de conectividade para definir a fragmentação de uma rede:

$$F = 1 - C_{on} = 2 * \frac{PNC}{N(N - 1)} \quad (2)$$

Quando $F = 0$, a rede é conectada, isto é, para cada par de vértices v_i e v_j existe um caminho conectando v_i a v_j . Quando $F = 1$, a rede é formada por vértices isolados, ou seja, para cada par de vértices v_i e v_j não existe um caminho conectando v_i a v_j .

Para cada rede de coautoria, obteve-se uma rede semântica baseada nos títulos publicados por seus coautores. Para que cada *RST* esteja conectada, foi necessária a exclusão de apenas um título do periódico *AFE*. A construção de cada *RST* obedeceu, resumidamente, os seguintes passos:

1) Os títulos dos periódicos passaram por um pré-tratamento que consistiu em aplicar uma série de regras tais como “nomes próprios devem formar uma única palavra”, “as sequências de palavras que têm um significado próprio, devem formar uma única palavra”, etc. Estas regras estão especificadas em [Fadigas 2011] e [Pereira et al. 2011].

2) Os títulos dos periódicos passaram por um tratamento computacional que consiste basicamente em eliminar palavras sem significados intrínsecos e em seguida, alterar as palavras restantes para a sua forma canônica. Para este tratamento utilizou-se o conjunto de programas desenvolvido por [Caldeira et al. 2006].

Após o tratamento, todas as palavras em um mesmo título estão conectadas entre si, formando um subgrafo completo ou uma clique. Assim, cada título gera uma clique. Duas cliques estarão conectadas se uma palavra ou mais pertencer(em) a dois títulos (cliques). Para cada periódico, formou-se uma rede semântica de cliques. A construção dessas redes podem ser encontradas com detalhes em [Fadigas 2011] e [Pereira et al. 2011].

Para cada *RST*, aplicou-se a estratégia de remoção uniforme e aleatória de vértices (*RR*). Realizaram-se 1000 retiradas para cada valor de fração de vértices removidos (f) listados a seguir: 0, 05; 0, 1; 0, 15; ...; 0, 65; 0, 7; 0, 71; 0, 72; ...; 0, 98; 0, 99. Logo, para cada rede de coautoria obtiveram-se 43.001 redes semânticas, totalizando 86.002 redes, sendo que 80.000 foram resultantes de remoção aleatória e uniforme de vértices.

Na etapa seguinte, observou-se o comportamento do tamanho da maior componente em função da fração de vértices removidos o que permitiu apontar a existência de um ponto crítico em $f = 0,7$. A maior componente é um subgrafo maximal conectado. Um subgrafo de uma rede é um grafo formado por um subconjunto do conjunto de

vértices da rede e por um subconjunto de arestas da rede. Uma rede cuja fragmentação é igual a zero possui apenas uma componente.

No ponto crítico $f = 0,7$, obtiveram-se redes semânticas uniformes nas quais identificaram-se os títulos que continham pelo menos uma palavra presente em cada rede. A partir dos títulos identificados nas redes semânticas uniformes, construíram-se redes de coautoria associadas. Cada rede de coautoria associada foi comparada com a sua respectiva rede de coautoria do periódico.

4. Resultados

As redes de coautoria dos periódicos *AFE* e *PEM* possuem, respectivamente, 994 e 951 vértices. Ambas as redes são altamente fragmentadas e o tamanho da maior componente não ultrapassa 16%. As redes semânticas dos periódicos *AFE* e *PEM* possuem, respectivamente, 1558 e 1186 vértices e ambas são conectadas. A Tabela 1 mostra dados gerais das redes de coautoria e das redes semânticas dos periódicos *AFE* e *PEM* e as Figuras 4a, 4b, 4c e 4d apresentam uma visualização de cada uma delas.

Tabela 1. Dados gerais para as redes de coautoria e redes semânticas baseadas em títulos de artigos dos periódicos *AFE* e *PEM*.

Índices	Coautoria		Semântica	
	<i>AFE</i>	<i>PEM</i>	<i>AFE</i>	<i>PEM</i>
Vértices	994	951	1558	1186
Arestas	1743	971	17877	12903
Densidade	0,00353	0,00215	0,01474	0,018362
Número de Componentes	230	296	1	1
Maior Componente (%)	3,01811	15,35226	100	100
Fragmentação	0,99404	0,97249	0	0
Número de Títulos	370	703	370	703
Número de Autores — Palavras	1214	1467	3984	5247
Autor por Vértice	1,22133	1,54259	—	—
Palavra por Vértice	—	—	2,55712	4,42411
Autores por Título	3,28108	2,08677	—	—
Palavras por Título	—	—	10,76757	7,46373
Títulos por Autor	0,30478	0,47921	—	—

A densidade de uma rede é um índice de coesão definida pela razão entre o seu número de arestas e o número possível de arestas na rede. A Tabela 1 mostra que todas as quatro redes possuem uma baixa densidade, caracterizando-as como redes esparsas. A rede de coautores do periódico *AFE* é mais coesa do que a rede de coautores do periódico *PEM*. Contudo, a rede semântica do periódico *PEM* é mais coesa do que o periódico *AFE*.

O número de vértices distingue-se do número de autores visto que, diferentemente do primeiro, o segundo pode ter repetições já que um autor pode publicar mais de um artigo. De forma análoga, o número de vértices distingue-se do número de palavras, pois uma palavra pode aparecer em mais de um título. O número médio de autores por vértice da rede de coautoria do periódico *AFE* é um pouco menor do que o da rede de

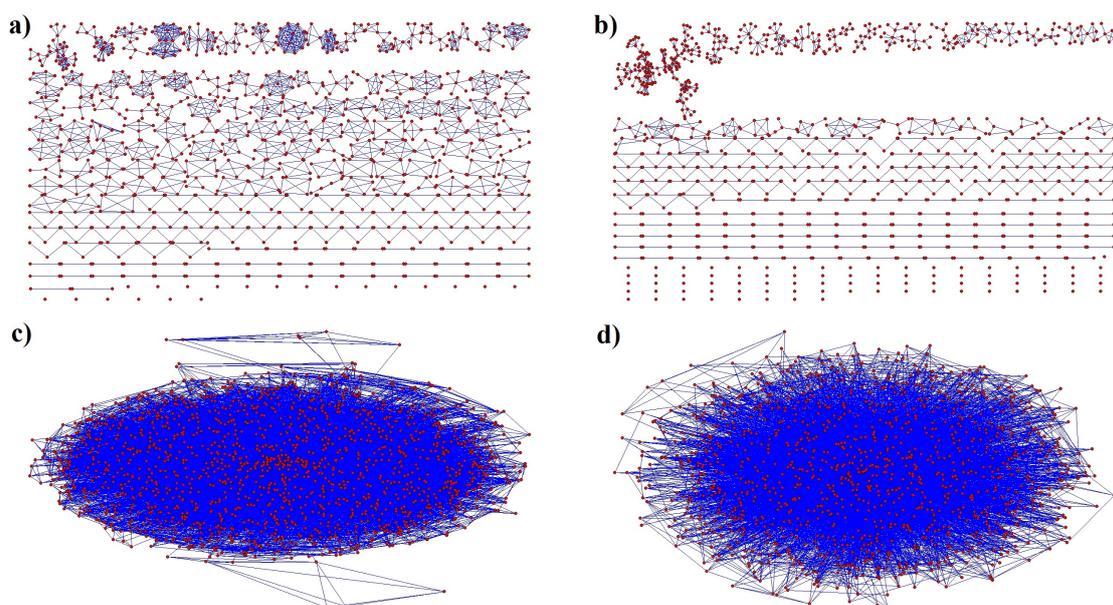


Figura 1. Visualização das redes de coautoria e das redes semânticas: a) periódico *AFE*; b) periódico *PEM* e c) periódico *AFE*; d) periódico *PEM*, respectivamente.

coautoria do periódico *PEM*, indicando que autores tendem a publicar novamente no periódico *PEM* com mais frequência do que no periódico *AFE*. Em se tratando das redes semânticas dos periódicos *AFE* e *PEM*, o número médio de palavras por vértice na primeira rede é menor do que na segunda. Isto indica que os autores do periódico *AFE* tendem a escolher mais palavras que ainda não apareceram nos títulos (do periódico *AFE*) do que no periódico *PEM*.

O número médio de autores por título no periódico *AFE* é maior do que no periódico *PEM*, o que é um reflexo da maior coesão da rede de coautoria do primeiro periódico. De outro modo, a produção em coautoria no periódico *AFE* tende a ser maior do que no periódico *PEM*. Os títulos no periódico *AFE* tendem a ter mais palavras do que no periódico *PEM*, o que pode contribuir para que os autores do periódico *AFE* utilizem mais palavras que ainda não apareceram em títulos de artigos já publicados (no periódico *AFE*). O número médio de títulos por autor no periódico *AFE* é menor do que no periódico *PEM*, indicando que no primeiro periódico a produção individual é mais baixa do que no segundo.

Para analisar as redes resultantes da remoção aleatória e uniforme de vértices das redes semânticas dos periódicos *AFE* e *PEM*, estimou-se o tamanho da maior componente. [Chen et al. 2007] usou a fragmentação (F) para mostrar que é possível estimar o tamanho da maior componente (P_∞) de uma rede após a remoção de vértices por estratégias *RR* e *HDR* (remoção de vértices de grau mais alto) para redes aleatórias e livre de escala e para valores de f menores que a fração crítica:

$$P_\infty \approx \sqrt{1 - F} \quad (3)$$

O valor de P_∞ é a razão entre o número de vértices da maior componente da rede após a remoção de vértices e o número de vértices da rede total conectada. A Figura 2a indica

que a relação $S \approx \sqrt{1 - F}$ também pode ser usada para estimar o tamanho da maior componente de redes semânticas de cliques, onde S é a razão entre o número de vértices da maior componente e o número de vértices da rede obtida após a remoção de vértices. A Figura 2b apresenta a média aritmética da estimativa do tamanho da maior componente S em função de f para as redes semânticas dos periódicos *AFE* e *PEM*. Cada valor de f no gráfico (exceto $f = 0$) é a média aritmética da estimativa da maior componente de 1000 redes. Para $f = 0$, tem-se a maior componente da rede conectada total.

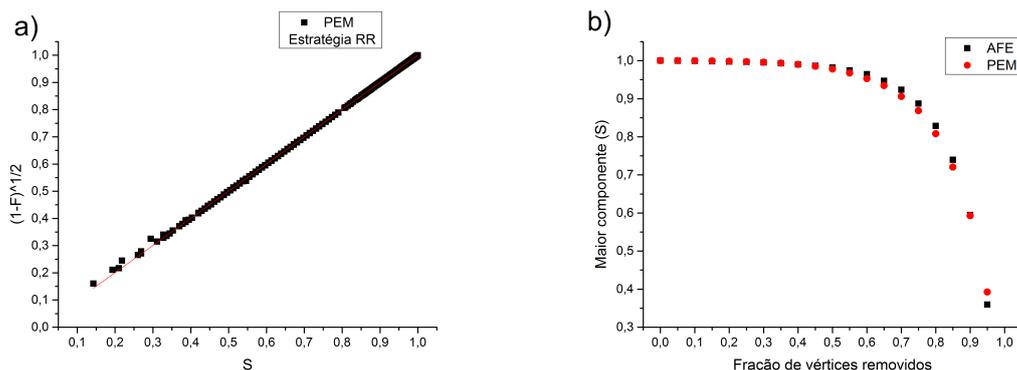


Figura 2. Estimativa do tamanho da maior componente das redes semânticas submetidas à remoção aleatória e uniforme de vértices: a) relação entre o tamanho da maior componente (S) do periódico *PEM* e a raiz quadrada de $1 - F$. Para cada fração de vértice removidos igual a 0, 1; 0, 3; 0, 5; 0, 7; 0, 9, comparou-se o tamanho da maior componente de 278 redes. O R^2 ajustado obtido foi igual a 0, 99993; b) média aritmética da estimativa do tamanho da maior componente (S) das *RST*'s dos periódicos *AFE* e *PEM* obtidas da remoção aleatória e uniforme de vértices.

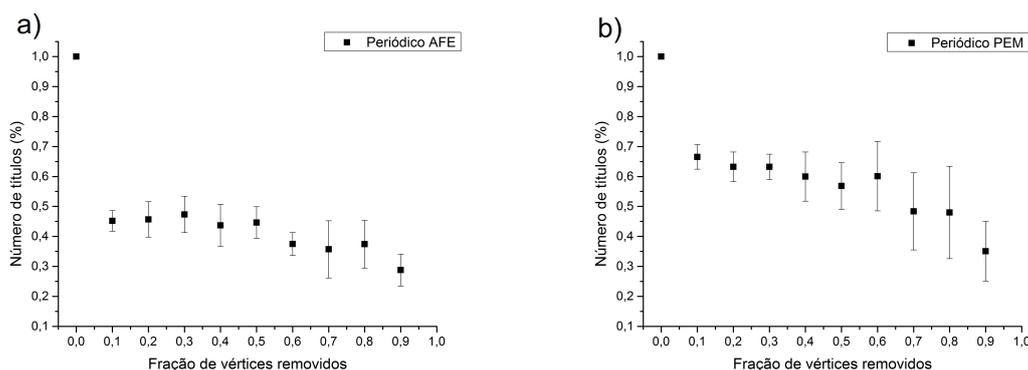


Figura 3. Média aritmética e desvio padrão da percentagem de títulos identificados para cada periódico em função da fração dos vértices removidos: a) periódico *AFE* e b) periódico *PEM*.

Os resultados encontrados indicam a existência de um ponto crítico para $f = 0, 7$ e neste ponto, tem-se redes semânticas uniformes. O tamanho da maior componente (S) de cada *RST* possui uma alta probabilidade de ser maior do que 90% para $f = 0, 7$ nos dois periódicos analisados. A alta resiliência sugere que para valores de f menores do que ou iguais a 0, 7 a rede mantém uma alta conectividade quando submetidas à estratégia

RR , ou seja, há uma tendência de existir uma alta conectividade entre as palavras. A média aritmética da conectividade de cada rede permanece acima de 0,8 para valores de f menores do que ou iguais a 0,7.

A partir de cada rede semântica resultante de remoção de vértices identificaram-se os títulos que contém pelo menos uma palavra presente na rede. Percebeu-se que a remoção de 10% de vértices implica que as palavras remanescentes na rede estão presentes em média em 45% e 67% dos títulos dos periódicos *AFE* e *PEM*, respectivamente. Contudo, a remoção de 70% de vértices implica que as palavras remanescentes na rede estão presentes em média em 36% e 48% dos títulos dos periódicos *AFE* e *PEM*, respectivamente. As Figuras 3a e 3b mostram a média aritmética e desvio padrão da percentagem de títulos identificados para cada periódico em função da fração dos vértices removidos.

A partir de redes semânticas uniformes ($f = 0,7$) construíram-se, para cada periódico, redes de coautoria associadas. A Tabela 2 mostra os índices das redes de coautoria associadas. As Figuras 4a, 4b, 4c e 4d mostram uma visualização das redes semânticas uniformes e das redes de coautoria associadas dos periódicos *AFE* e *PEM*

Tabela 2. Índices das redes de coautoria associadas para os periódicos *AFE* e *PEM*.

Índices	<i>AFE</i>	<i>PEM</i>
Vértices	391,1	505,9
Arestas	656,2	487,2
Densidade	0,00905	0,00395
Títulos	132,1	340
Autores	450,3	715
Autores por vértice	1,15137	1,41332
Autores por título	3,40878	2,10294
Títulos por autor	0,29336	0,47552

Os índices apresentados pela Tabela 2 evidenciam que uma rede de coautoria associada tende a manter o mesmo comportamento da sua rede de coautoria do periódico (Tabela 1). Nota-se, por exemplo, que a rede de coautoria associada do periódico *AFE* tende a ser mais coesa do que a rede de coautoria associada do periódico *PEM*, mantendo o que já ocorre com as redes de coautoria dos periódicos *AFE* e *PEM*. As redes de coautoria associadas do periódico *AFE* tendem a ter um número médio de autores por vértice menor do que os das redes de coautoria associadas do periódico *PEM*. Esta constatação coincide com a análise feita anteriormente a partir dos dados apresentados na Tabela 1 de que autores tendem a publicar novamente no periódico *PEM* com mais frequência do que no periódico *AFE*.

O número de médio de autores por título tende a ser maior em uma rede de coautoria associada do periódico *AFE* do que em uma rede de coautoria associada do periódico *PEM*. Situação semelhante acontece com as redes de coautoria dos periódicos *AFE* e *PEM*, conforme Tabela 1. O número de médio de títulos por autor das redes de coautoria associadas do periódico *AFE* também tende a ser menor do que o número médio de títulos por autor das redes de coautoria associadas do periódico *PEM*, refletindo uma

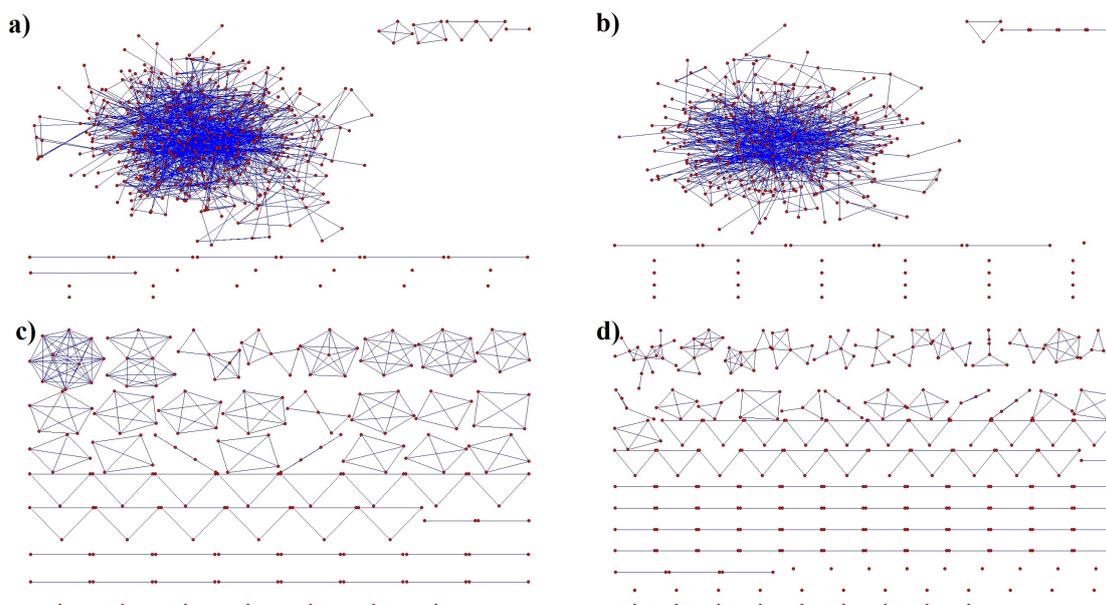


Figura 4. Visualização das redes semânticas uniformes e das redes de coautoria associadas: a) periódico AFE; b) periódico PEM e c) periódico AFE; d) periódico PEM, respectivamente.

menor produção individual no primeiro periódico do que no segundo. A comparação dos dados das Tabelas 1 e 2 aponta evidências de que há uma integração entre coautores quando estes lidam com a escolha de palavras para os títulos de seus artigos, considerando a identidade do periódico. A Tabela 3 mostra a lista de 10 autores com maior grau na rede de coautoria e em uma rede de coautoria associada do periódico PEM. O grau (k_i) de um vértice v_i é o número de arestas incidentes no vértice.

Tabela 3. Lista de 10 autores com maior grau na rede de coautoria e em uma rede de coautoria associada do periódico PEM.

Ordem	Coautoria original - PEM		Coautoria associada - PEM	
	Autor	k_i	Autor	k_i
1	yklin	18	yklin	9
2	bfspencerjr	14	mohammadnoori	6
3	wqzhu	11	ykwen	6
4	gischueller	11	georgedeodatis	6
5	mrigoriu	11	yukiofujimoto	6
6	oveditlevsen	9	seiichiito	6
7	georgedeodatis	9	masanobushinozuka	6
8	gqcai	9	mnoori	6
9	mfdimentberg	9	zhou	6
10	cpapadimitriou	9	jamestpyao	5

Nesta lista, observou-se que 30% dos autores com grau elevado nas redes de coautoria dos dois periódicos também tendem a ter grau elevado nas redes de coautoria associadas. Desta forma, há indícios de coerência entre 30% dos autores nas redes asso-

ciadas indicando que a rede semântica uniforme tende a manter uma estrutura próxima da rede semântica original, antes da remoção de vértices.

5. Considerações Finais

Este trabalho analisou redes de coautoria de um mesmo periódico por meio de redes semânticas uniformes. O uso de redes semânticas baseadas em títulos de artigos científicos pode auxiliar no entendimento da integração de coautores de um mesmo periódico. Desta forma, pode-se apontar evidências de como autores decidem os títulos de seus artigos científicos.

Redes semânticas baseadas em títulos de periódicos científicos tendem a ter um ponto crítico quando se retira aleatoriamente e uniformemente 70% dos vértices. A existência de um ponto crítico para *RSTs* indica a possibilidade de se poder representar uma rede semântica por meio de uma rede semântica uniforme. Desta forma, o método proposto neste trabalho pode auxiliar pesquisadores na busca de temas relacionados ao seu objeto de investigação.

Referências

- Bales, M. E. and Johnson, S. B. (2006). Graph theoretic modeling of large-scale semantic networks. *Journal of Biomedical Informatics*, 39(4):451–464.
- Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.
- Borgatti, S. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, 12(1):21–34.
- Caldeira, S. M. G., Lobão, T. C. P., Andrade, R. F. S., Neme, A., and Miranda, J. G. V. (2006). The network of concepts in written texts. *The European Physical Journal B - Condensed Matter and Complex Systems*, 49(4):523–529.
- Cancho, R. F. and Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482):2261–2265.
- Carvalho, D. B. F., Marcacini, R., Lucena, C., and Rezende, S. (2012). Towards a process to support solving the content selection problem from online community forums. In *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining, XXXII Congress of the Brazilian Computer Society Computer Society*.
- Chen, C., Song, I.-Y., Yuan, X., and Zhang, J. (2008). The thematic and citation landscape of data and knowledge engineering (1985-2007). *Data and Knowledge Engineering*, 67(2):234–259.
- Chen, Y., Paul, G., Cohen, R., Havlin, S., Borgatti, S. P., Liljeros, F., and Stanley, H. E. (2007). Percolation theory and fragmentation measures in social networks. *Physica A: Statistical Mechanics and its Applications*, 378(1):11–19.
- Fadigas, I. S. (2011). *Difusão do Conhecimento em Educação Matemática sob a perspectiva das Redes Sociais e Complexas*. PhD thesis, Universidade Federal da Bahia.
- Fadigas, I. S. and Pereira, H. B. B. (2013). A network approach based on cliques. *Physica A*, 392(10):2576–2587.

- Katz, J. S. and Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26:1–18.
- Krackhardt, D. (1994). *Graph Theoretical Dimensions of Informal Organizations in: Computational Organizational Theory*. Lawrence Erlbaum Associates.
- Laender, A. H. F., Moro, M. M., Silva, A. S., Davis Jr., C. A., Gonçalves, M. A., Galante, R., Silva, A. J. C., Bigonha, C. A. S., Dalip, D. H., Barbosa, E. M., Borges, E. N., Cortez, E., Procópio Jr., P., Alencar, R. O., Cardoso, T. N. C., and Salles, T. (2011). Ciênciabrasil - the brazilian portal of science and technology. In *Seminário Integrado de Software e Hardware, 38, Natal. Anais eletrônicos... Natal: Seminário Integrado de Software e Hardware*.
- Lopes, G. R., Moro, M., Silva, R. d., Barbosa, E. M., and Oliveira, J. P. M. (2011). Ranking strategy for graduate programs evaluation. In *ICITA - 7th International Conference on Information Technology and Application, Sydney, Austrália*.
- Newman, M. E. J. (2001). Scientific collaboration networks. i. network construction and fundamental results. *Physical Review E*, 64.
- Pereira, H. B. B., Fadigas, I. S., Senna, V., and Moret, M. A. (2011). Semantic networks based on titles of scientific papers. *Physica A: Statistical Mechanics and its Applications*, 390(6):1192–1197.
- Quintella, R. H., Freitas, E. J. d. S. M. d., Ventura, A. C., Santos, M. A., and Antonio, L. Q. (2009). Network dynamics in scientific knowledge acquisition: an analysis in three public universities in the state of bahia. *Revista de Administração Pública*, 43(6):1279–1314.
- Rosa, M. G., Fadigas, I. d. S., Andrade, M. T. T., and Pereira, H. B. B. (2012). Abordagem de redes por cliques: aplicação a redes de coautoria. In *In: BraSNAM - Brazilian Workshop on Social Network Analysis and Mining, Anais do XXXII Congresso da Sociedade Brasileira de Computação*, volume 1.
- Teixeira, G. M., Aguiar, M. S., Carvalho, C. F., Dantas, D. R., Cunha, M. V., Morais, J. H. M., Pereira, H. B. B., and Miranda, J. G. V. (2010). International journal modern physics c. *Complex Semantic Network*, 21(3):333–347.
- Vanz, S. A. S. and Stump, I. R. C. (2010). Colaboração científica: revisão teórico conceitual. *Perspectivas em Ciência da Informação*, 15(2):42–55.
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 6684(393):440–442.