

# Análise de redes semânticas baseada em títulos de artigos de periódicos científicos: o caso dos periódicos de divulgação em educação matemática

INÁCIO DE SOUSA FADIGAS\*

TRAZÍBULO HENRIQUE\*\*

VALTER DE SENNA\*\*\*

MARCELO A. MORET\*\*\*\*

HERNANE BORGES DE BARROS PEREIRA\*\*\*\*\*

## Resumo

Redes sociais e complexas representam uma das formas de se estudarem sistemas complexos. A proposta do presente artigo é descrever e analisar redes semânticas construídas a partir dos títulos de artigos de periódicos nacionais de divulgação da educação matemática. A escolha dos periódicos considerou, principalmente, um público mais ampliado, composto por estudantes e professores de matemática. Para realizar o estudo, a análise de redes sociais e a teoria das redes complexas foram usadas. O artigo descreve ainda o método para construção das redes e os resultados são apresentados em três blocos: os quantitativos gerais, os índices da teoria das redes complexas e os índices da análise de redes sociais. Finalmente, é realizado um diagnóstico sobre a caracterização das redes semânticas, assim como sobre a importância das palavras e sua frequência nos títulos, semelhanças na linha editorial, vocabulário padrão, entre outras constatações. Uma importante contribuição deste trabalho é de oferecer suporte para a definição de estratégias que captem mais leitores, ajudando no processo de difusão do conhecimento em campos específicos.

*Palavras-chave:* análise de redes sociais; redes complexas; educação matemática.

---

\* Departamento de Ciências Exatas – Universidade Estadual de Feira de Santana. E-mail: isfadigas@gmail.com

\*\* Departamento de Ciências Exatas – Universidade Estadual de Feira de Santana. E-mail: henrique@uefs.br

\*\*\* Programa de Modelagem Computacional – Senai – Cimatec. E-mail: vsenna@terra.com.br

\*\*\*\* Programa de Modelagem Computacional – Senai – Cimatec. Departamento de Física – Universidade Estadual de Feira de Santana. E-mail: mamoret@gmail.com

\*\*\*\*\* Programa de Modelagem Computacional – Senai – Cimatec, Departamento de Ciências Exatas – Universidade Estadual de Feira de Santana. E-mail: hbbpereira@gmail.com

### **Abstract**

*Social and complex networks provide an important contribution to the understanding and representation of complex systems. The purpose hereof is to describe and analyze semantic networks built on titles of papers that were published in Brazilian journals on mathematics education. The journals of choice took mainly into account the site of the audience, composed of students and math teachers. In order to carry out the proposed study, both social network analysis and complex network theory were used. The paper also describes the method for constructing semantic networks and findings are presented in three groups: quantitative results, indices derived from the complex network theory, and indices from social network analysis. Finally, a diagnosis on the semantic network characterization is performed, as well as on the importance of words and their frequency in the titles of papers, similarities in the editorial line and standard vocabulary, among other findings. An important contribution hereof is to offer support for developing strategies to entice more readers, helping in the process of the diffusion of knowledge on specific areas.*

**Keywords:** *social network analysis; complex networks; mathematics education.*

### **Introdução**

O interesse de pesquisadores em analisar atores (e.g. indivíduos, organizações, páginas *web*, computadores, palavras, neurônios, etc.) e seus relacionamentos, cujo fundamento matemático se assenta na teoria dos grafos, vem promovendo, desde o início do século XX, descobertas importantes sobre como redes de relacionamento são estabelecidas e funcionam. Com o advento, o avanço e a popularização dos computadores a partir da segunda metade do século XX, muitas redes que não eram consideradas, devido à grande quantidade de atores e relacionamentos, passam a ser foco de estudo. Cientistas das mais diferentes áreas do conhecimento mostram seu interesse por redes em geral e investem muito esforço no estudo de propriedades emergentes. Surge uma área de pesquisa multidisciplinar cujos alicerces estão fincados na física: redes complexas.

Nesse contexto científico, constata-se um avanço teórico-tecnológico contínuo na pesquisa e no desenvolvimento de técnicas e métodos para trabalhar com redes sociais e complexas. Essas técnicas e esses métodos de análise podem ser aplicados a diferentes contextos (e.g. amizade, desenvolvimento de pesquisa, relacionamentos empresariais, relação entre palavras em um discurso, etc.), de modo a modelar as relações entre os atores observados e explicitar/explicar a estrutura de uma determinada rede.

Considerando isso, o contexto desta pesquisa foca a análise dos relacionamentos entre palavras. Colocamos nossas atenções no estudo de redes semânticas baseadas em títulos de artigos publicados em periódicos

científicos. Para tanto, respaldamo-nos em três teorias: teoria dos grafos, análise de redes sociais e teoria das redes complexas.

A teoria dos grafos, a análise de redes sociais e a teoria das redes complexas têm sido usadas para investigar o comportamento e as estruturas de diversos tipos de redes, e.g. redes tecnológicas (Watts e Strogatz, 1998; Barabási e Albert, 1999; Faloutsos et al., 1999); redes biológicas (Jeong et al., 2000; Liljeros et al., 2001); redes econômicas (Pereira et al., 2007); redes de informação (Newman, 2001a, 2001b); redes de disseminação de enfermidades (Vieira, 2005) e redes semânticas (Ferrer i Cancho e Solé, 2001; Caldeira et al., 2006).

Neste artigo, apresentamos uma discussão sobre a divulgação científica em educação matemática a partir de um diagnóstico fundamentado em redes semânticas. Esse tipo de rede representa um dos fundamentos teóricos do estudo de processos cognitivos, uma vez que pode ser usado para representar ou inferir sobre conhecimento. A fonte principal dos dados usados para a análise é o conjunto de títulos dos artigos científicos e/ou de divulgação publicados nos periódicos nacionais cujos focos estão dirigidos a estudantes, professores de ensino fundamental, médio e superior e o público em geral, que tenham interesse em educação matemática: *Boletim GEPEN*, *Revista do Professor de Matemática (RPM)*, *Boletim de Educação Matemática (BOLEMA)*, *Educação Matemática em Revista (EMR)*, *Folbetim de Educação Matemática (Folbetim)* e *Zetetiké*. Assim, neste estudo, dizemos que uma rede semântica é originada na associação de palavras de títulos de artigos científicos. Os atores ou vértices são as palavras selecionadas por seus significados intrínsecos e os relacionamentos ou arestas ocorrem entre todas as palavras de um título.

Para realizar este estudo, utilizamos um método de construção de redes semânticas que consiste basicamente na (1) eliminação das palavras sem significados intrínsecos e na (2) alteração das palavras restantes para sua forma canônica. Cada título forma uma rede na qual todos os vértices (i.e. palavras) estão conectados entre si, formando assim cliques ou subgrafos completos. Dessa forma, construímos uma rede semântica baseada nos títulos de artigos publicados em um periódico específico.

Ressaltamos que a análise de redes semânticas de títulos em educação matemática é uma tentativa de entender a associação entre os textos científicos e de divulgação e seus títulos, de modo a verificar as influências que os títulos sofrem com respeito (1) ao jargão e aos termos técnicos

da área ou do periódico científico, (2) às pesquisas desenvolvidas em um determinado período e (3) a personalidades importantes que passam a ser o foco principal em determinados artigos. Ademais, a análise realizada permite refletir sobre o impacto que a estrutura de um título pode causar do ponto de vista da difusão do conhecimento.

O artigo está organizado como segue. A Seção 2 detalha os procedimentos metodológicos estabelecidos para construir as redes semânticas propostas. A Seção 3 estende a anterior, apresentando os resultados e a reflexão da análise das redes semânticas sob alguns aspectos, principalmente sobre sua contribuição com respeito à difusão do conhecimento. A Seção 4 apresenta as conclusões da pesquisa de que trata este artigo.

## 1. Método para construir as redes semânticas

### 1.1. Critério de escolha dos periódicos

O principal requisito para a escolha dos periódicos está fortemente ligado ao caráter de difusão da matemática para um público de não especialistas. Não se pretendeu nesta pesquisa incluir os vários veículos de divulgação científica em matemática disponíveis no país, voltados, sobretudo para pesquisadores. O foco foi dirigido para aqueles que se voltam para estudantes, professores de ensino fundamental, médio e superior e o público em geral que tenham interesse em educação matemática. A seguir, apresentam-se os propósitos dos periódicos segundo seus próprios textos:

- *Boletim GEPEM (GEPEM)*: Veículo de divulgação do Grupo de Estudos e Pesquisas em Educação Matemática. Publica seus boletins desde 1976.

Na procura constante de aperfeiçoamento do ensino de matemática, (...) de congregar e interessar colegas de matemática ou de outras disciplinas em torno desse objetivo. (n° 2, abril de 1977)

- *Revista do Professor de Matemática (RPM)*: Veículo de divulgação da Sociedade Brasileira de Matemática (SBM) que publica a revista desde 1982.

A *Revista do Professor de Matemática* é uma publicação semestral da Sociedade Brasileira de Matemática. O escopo da revista é o de se constituir num ponto de encontro de professores de matemática atuantes no 1° e 2° graus, contando experiências, procurando

respostas, discutindo sugestões, divulgando notícias. (n° 1, 2° semestre de 1982)

- *Boletim de Educação Matemática (BOLEMA)*: Publicado pelo Programa de Pós-Graduação em Educação Matemática da UNESP desde 1985.

O *BOLEMA* pretende ser um veículo informal de divulgação dos fatos e notícias relacionados com a Educação Matemática e um meio rápido de comunicação entre os professores de Matemática. (Ano 1. n. 1, 1985)

- *Educação Matemática em Revista (EMR)*: Publicado pela Sociedade Brasileira de Educação Matemática (SBEM) desde 1993.

Ao elaborarmos o projeto desta Revista, (...) tínhamos em mente de que a Revista pretendia, “através do conhecimento e da interação contínuos com o ensino e com o professor e conferindo a este um intenso papel na participação, tornar-se um locus de expressão, discussão e busca de soluções para os problemas do ensino de matemática de 1° e 2° graus, nos cursos de Magistério e de Licenciatura em Matemática e constituir-se em fonte coletiva para a melhoria desse ensino.” (Ano I – n. 2, 1994)

- *Folhetim de Educação Matemática (Folhetim)*: Divulgado pelo NEMOC – Núcleo de Educação Matemática Omar Catunda da Universidade Estadual de Feira de Santana, desde 1993.

Este Folhetim é um veículo de divulgação, circulação de idéias e de estímulo ao estudante e à curiosidade intelectual. Dirige-se a todos os interessados pelos aspectos pedagógicos, filosóficos e históricos da Matemática. Pretende construir uma ponte para unir os que estão próximos e os que estão distantes. (Ano 1, n. 1, 1993)

- *Zetetiké* é uma revista da Faculdade de Educação da Universidade Estadual de Campinas (Unicamp), cujo primeiro número data de 1993. Está ligada ao Centro de Estudo, Memória e Pesquisa em Educação Matemática (Cempem) que busca entre outros objetivos:

[...] constituir-se em um veículo de interação científico-pedagógico entre pesquisadores e educadores matemáticos que atuam no ensino de 1° e 2° graus. (Ano 1, n. 1, 1993)

Pelo exposto, observa-se que os periódicos supracitados têm uma preocupação central com ensino de matemática, atualmente envolvido por um espectro mais amplo que é a Educação Matemática.

## 1.2. Delimitação dos dados

Considerando o caráter dinâmico da pesquisa, uma vez que, a cada ano, de 1 a 4 números dos periódicos supracitados são publicados, os dados coletados abrangem um período do primeiro número até os indicados na Tabela 1.

Tabela 1 – Relação dos periódicos científicos nacionais usados na obtenção de dados para a criação das redes semânticas dos títulos dos artigos selecionados

#	Título	Referência		Qt. Títulos
		Período	Números	
1	<i>GEPEM</i>	1976 a 2008	1 a 52	262
2	<i>RPM</i>	1982 a 2006	1 a 61	580
3	<i>BOLEMA</i>	1985 a 2008	1 a 29	196
4	<i>EMR</i>	1993 a 2007	1 a 22	148
5	<i>Folbetim</i>	1993 a 2007	1 a 141	178
6	<i>Zetetiké</i>	1993 a 2006	1 a 26	119

Fonte: Autores.

O Quadro 1 traz outras informações sobre os periódicos, no que diz respeito à classificação em relação a QUALIS da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Note-se que este não se constituiu em um critério para a escolha dos periódicos, mas pode-se notar que, com a exceção do *Folbetim de Educação Matemática*, todos eles são classificados, sendo que 4 deles o são com a classificação máxima A e circulação nacional, na área Ensino de Ciências e Matemática.

Quadro 1 – Classificação dos periódicos científicos nacionais que publicam artigos relacionados principalmente com Educação Matemática

ISSN	Título	Qualis*	Circulação	Área de avaliação
0104-9739	GEPEN	A	Nacional	ENSINO DE CIÊNCIAS E MATEMÁTICA
		B	Nacional	EDUCAÇÃO
0102-4981	RPM	C	Nacional	ENSINO DE CIÊNCIAS E MATEMÁTICA
		C	Nacional	EDUCAÇÃO
0103-636X	BOLEMA	A	Nacional	ENSINO DE CIÊNCIAS E MATEMÁTICA
		B	Nacional	EDUCAÇÃO
		B	Nacional	INTERDISCIPLINAR
1517-3941	EMR	A	Nacional	ENSINO DE CIÊNCIAS E MATEMÁTICA
		A	Nacional	EDUCAÇÃO
1415-8779	Folbetim	Sem classificação		
0104-4877	Zetetiké	A	Nacional	ENSINO DE CIÊNCIAS E MATEMÁTICA
		B	Nacional	EDUCAÇÃO

\* A classificação foi obtida da página QUALIS da Capes (<http://qualis.capes.gov.br/web-qualis/>), acessada em 25 de outubro de 2008.

Fonte: Autores.

### 1.3. Tratamento dos textos (manual e com o uso de pacotes)

Para a construção da rede de títulos, dois procedimentos principais foram necessários: o tratamento manual e o tratamento com o uso de programas computacionais. Por tratamento manual entende-se a fase inicial de formatação dos dados. Esta constou da disposição dos títulos em um arquivo de texto, cada título escrito em uma única linha. Em seguida, fez-se necessário aplicar regras gerais preestabelecidas:

- a) os títulos em língua estrangeira foram traduzidos para o português;
- b) os nomes próprios devem formar uma única palavra (e.g. Rio de Janeiro);
- c) palavras repetidas no mesmo título são excluídas, restando apenas uma ocorrência;
- d) as seqüências de palavras que têm um significado próprio, devem formar uma única palavra (e.g. educaçãomatemática).

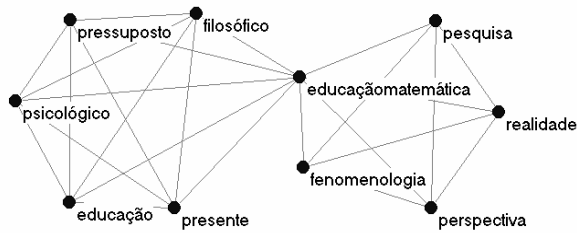
Para o tratamento com o uso de programas computacionais, a base foi o uso do conjunto de programas constantes do pacote UNITEX, disponibilizado pela Rede Relex Brasil, que faz parte de um programa francês do Laboratoire d'Automatique Documentaire et Linguistique (LADL). Além deste, foi usado o programa Ambisin, desenvolvido por Caldeira (2005) para tratar questões como a eliminação das ambigüidades, eliminação das palavras gramaticais e separação das formas flexionadas ou canônicas das palavras do restante dos itens de classificação gramatical gerados pelo UNITEX. Para a presente pesquisa, foi selecionada a opção do Ambisin que reduz palavras à forma canônica e elimina as palavras gramaticais (e.g. artigos, pronomes pessoais e possessivos, adjetivos possessivos, demonstrativos, interrogativos, advérbio, etc.). Para a descrição completa do método sugere-se o trabalho de Caldeira (2005).

#### **1.4. Construção das redes**

Após o tratamento com os programas, cada rede foi construída usando o programa NetPal (ibid.). As palavras (após os tratamentos) que ocorrem em um título estão ligadas entre si. Palavras de títulos distintos estão ligadas se ocorrem em ambos. A Figura 1 ilustra a formação da rede.

A Figura 1 apresenta a rede resultante de dois títulos (i.e. “Educação na Educação Matemática: os Pressupostos Filosóficos e Psicológicos de Educação Presentes na Educação Matemática” – T01 e “A pesquisa em Educação Matemática: Realidade e Perspectiva. A Fenomenologia” – T02) depois de aplicado o método descrito anteriormente. É possível observar que a palavra “educaçõamatemática” é o vértice de ligação entre os títulos. Quando toda a rede é finalmente construída, vértices desse tipo têm muito mais probabilidade de se tornarem pontos centrais de interesse.





T01: Educação na Educação Matemática: os Pressupostos Filosóficos e Psicológicos de Educação Presentes na Educação Matemática.

T02: A pesquisa em Educação Matemática: Realidade e Perspectiva. A Fenomenologia.

educação	N	pesquisa	N
eduçãomatemática	NOTFOUND	eduçãomatemática	NOTFOUND
pressuposto	N	realidade	N
filosófico	A	perspectiva	N
psicológico	A	fenomenologia	N
presente	N		

Figura 1 – Exemplo de formação da rede de palavras dos títulos. Após o tratamento de um título de artigo com os programas UNITEX e Ambisin, subgrafos completos são construídos até a formação da rede contextualizada de títulos para aquele periódico. Depois de tratados os arquivos, um novo arquivo em formato texto é gerado com as palavras que serão usadas para construir a rede, dispostas linha a linha com a identificação de sua categoria (e.g. N para substantivo, A para adjetivo, NOTFOUND para palavras não encontradas no dicionário usado). Os títulos apresentados são do *Bolema*, número 4.

## 2. Resultados e discussões

As redes obtidas foram estudadas com o emprego de dois programas específicos para tal finalidade: o Pajek<sup>1</sup> e o Ucinet (Borgatti et al., 2002). A análise e discussão dos resultados foram realizadas considerando três perspectivas: índices ou quantitativos gerais, abordagem das redes complexas e abordagem das redes sociais.

### 2.1. Quantitativos gerais

As informações quantitativas sobre as redes podem ser resumidas na Tabela 2.

1 Programa para criação, análise e visualização de redes desenvolvido por Vladimir Batagelj and Andrej Mrvar – <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

O número de vértices expressa quantas palavras distintas (após o tratamento do texto) resultaram, ou seja, o vocabulário usado nos títulos de cada periódico. A relação entre o número de palavras e o número de vértices expressa o quanto em média as palavras se repetiram nos títulos (Tabela 2). Índices maiores apontam para uma diversidade maior nas palavras usadas nos títulos, ou seja, um menor número de palavras em comum. Já a relação entre palavras e títulos mede o comprimento médio para cada título, após o tratamento.

Tabela 2 – Índices das redes de palavras dos títulos

Índices	<i>GEPEM</i>	<i>RPM</i>	<i>BOLEMA</i>	<i>EMR</i>	<i>Folbetim</i>	<i>Zetetiké</i>
Número de vértices ( $n$ )	593	871	496	380	548	393
Número de arestas	2800	2222	2721	1970	4687	1966
Número de títulos ( $T$ )	262	580	196	148	178	119
Número de palavras	1282	1799	1109	817	1178	745
Componentes	14	67	4	5	11	5
Porcentagem do maior componente	93,4	83,7	98,8	97,9	97,3	97,2
Pontos de corte ( $P_c$ )	23	143	9	9	14	7
Densidade da rede	0,01595	0,00587	0,02216	0,02735	0,03127	0,02552
Relação no. de palavras/ no. de vértices	2,1618	2,0654	2,2358	2,1500	2,1496	1,8956
Relação no. de palavras/ no. de títulos	4,8931	3,1017	5,6581	5,5202	6,6179	6,2605
Relação títulos/edição do periódico	5,0	9,5	6,8	6,7	3,9	4,6

Fonte: Autores

Os resultados são apresentados na Tabela 2. Nas redes analisadas, o *Folbetim* apresenta, em média, títulos com muitas palavras, mais que o dobro da *RPM*, que se caracteriza por apresentar títulos curtos. Note que a relação palavras/títulos e a densidade da rede apresentam uma forte associação, destoando do conjunto apenas os índices da *EMR* (Figura 2). A propósito, a densidade da rede mede a porção de arestas presentes na rede, em relação à quantidade total possível, que é dada por  $n(n-1)/2$  onde  $n$  o número de vértices. A densidade maior de um periódico reflete uma maior quantidade de vínculos entre as cliques.

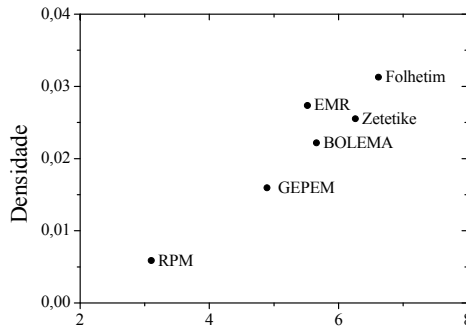


Figura 2 – Relação entre a média de palavras por títulos e a densidade da rede

O número de componentes de uma rede diz respeito à conectividade daquela rede. Uma rede com um único componente é aquela em que todos os vértices são ligados a cada um dos outros vértices e no caso da rede em estudo, significa que não existiriam títulos isolados. Porém, com a quantidade dos dados coletados, todas as redes têm mais de um componente. O aspecto visual é apresentado na Figura 3.

O número de componentes revela a diversidade semântica na escolha das palavras para compor os títulos. Nessa análise, destacam-se os títulos da *RMP*, que se distancia das outras publicações, deixando patente essa diversidade. Não há uma preocupação de convergência das palavras dos títulos para traduzir alguma linha editorial da revista. Um índice que afeta diretamente a quantidade de componentes é a média de palavras por título, mas este não é o único fator. Quanto menor esta média, menor a probabilidade de conexão entre dois títulos. A Figura 4 mostra este comportamento.

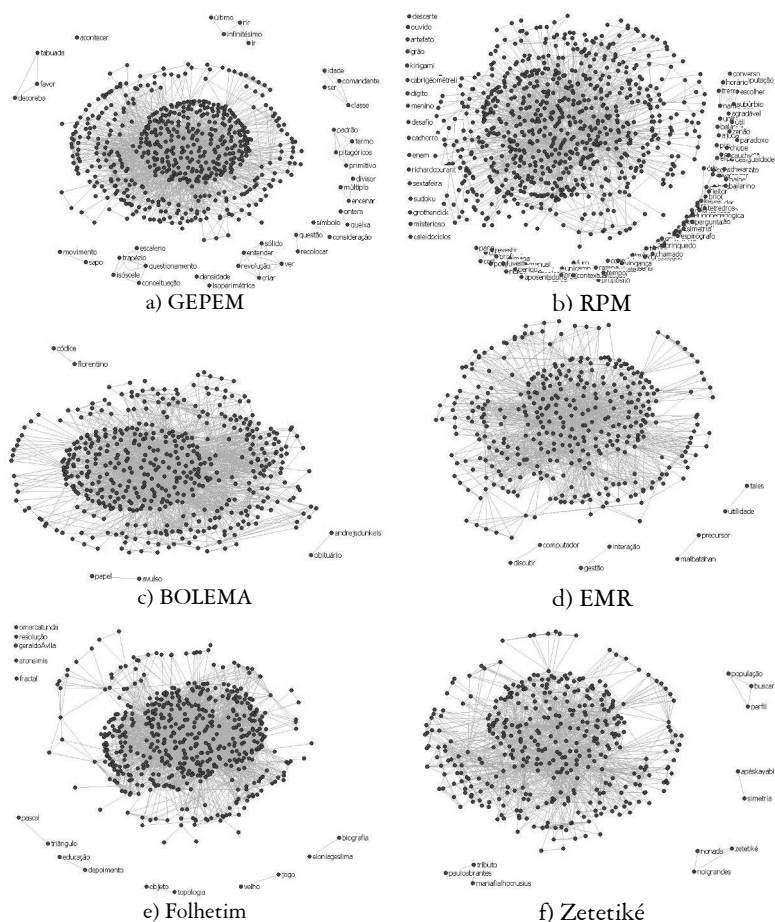


Figura 3. Componentes das redes: a) *GEPEM* – 14 componentes. Principal com 554 vértices; b) *RPM* – 67 componentes. Principal com 728 vértices; c) *BOLEMA* – 4 componentes. Principal com 490 vértices. d) *EMR* – 5 componentes. Principal com 372 vértices; e) *Folhetim* – 11 componentes. Principal com 533 vértices; f) *Zetetiké* – 5 componentes. Principal com 382 vértices. A disposição topológica obtida no balanceamento entre a repulsão dos vértices e o viés de igual comprimento das arestas (Ucinet).

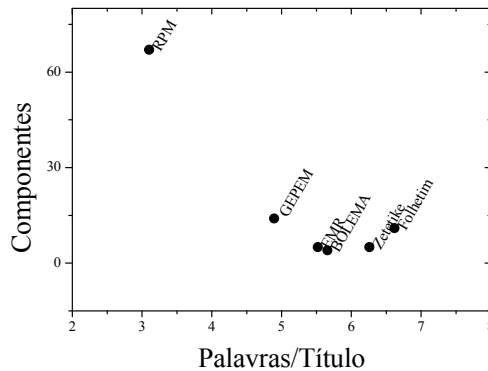


Figura 4 – Variação da média de palavras por título com a quantidade de componentes

## 2.2. Índices da teoria das redes complexas

Os resultados mais importantes quando se trata de redes complexas são: o coeficiente de aglomeração médio  $C_{ws}$ , o caminho mínimo médio  $L$  e a distribuição dos graus  $P(k)$ . Através do conhecimento de tais índices é possível caracterizar uma rede como “aleatória”, “livre de escala” ou “small world”. O coeficiente de aglomeração médio usado é o proposto por Watts e Strogatz (1998), que descreve a vizinhança imediata de um vértice  $i$  em uma rede. Define-se o coeficiente de aglomeração local  $C_i$  do vértice  $i$  como a fração dos pares de vértices vizinhos ao vértice  $i$  que estão eles mesmos conectados. O coeficiente de aglomeração da rede é a média dessa quantidade sobre todos os  $n$  vértices (Newman et al., 2006).

$$C_{ws} = \frac{1}{n} \sum_{i=1}^n C_i = \frac{1}{n} \sum_{i=1}^n \frac{\text{número de pares de vizinhos conectados}}{\frac{1}{2} k_i (k_i - 1)} \quad (01)$$

onde  $k_i$  é o grau do vértice  $i$ .

O caminho mínimo médio é a média das menores distâncias vértice a vértice, e é calculado por

$$\ell = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \quad (02)$$

onde  $d_{ij}$  é a menor distância do vértice  $i$  ao vértice  $j$  e  $n$  é o número de vértices.

Embora o coeficiente de aglomeração e o caminho mínimo médio possam ser calculados para redes desconectadas (i.e. aquelas com mais de

um componente), a rigor, devem surgir nos cálculos distâncias infinitas quando dois vértices não estão ligados. Portanto, para o que segue, usamos o maior componente da rede para a análise. O número de vértices presentes em tais componentes é alto, sendo que nos periódicos as percentagens deste índice são maiores que 93%, exceto em 1 deles, que apresenta uma percentagem de 83,7% (Tabela 2 e Tabela 3).

A Tabela 3 apresenta os resultados dos cálculos do caminho mínimo médio e do coeficiente de aglomeração médio, além de outros resultados para as redes conectadas.

Tabela 3 – Índices da análise de redes complexas para as redes conectadas

	GEPEM	RPM	BOLEMA	EMR	Folhetim	Zetetiké
Número de vértices ( $n$ )	554	728	490	372	533	382
Número de títulos	247	510	193	144	168	114
Coeficiente de aglomeração médio ( $C_{ws}$ )	0,8109	0,6528	0,8028	0,8137	0,8214	0,8453
Coeficiente de aglomeração médio (rede aleatória) $C_{ra} = z/n$	0,0179	0,0079	0,0226	0,0284	0,0329	0,0268
Caminho mínimo médio $\ell$	2,713	3,704	2,680	2,672	2,454	2,599
Caminho mínimo médio (rede aleatória) $\ell_{ra}$	2,962	4,011	2,858	2,754	2,497	2,810
Diâmetro	6	9	5	5	6	5
Grau médio ( $z$ )	9,938	5,819	11,094	10,570	17,568	10,246
$C_{ws} / C_{ra}$	45,20	81,67	35,52	28,65	24,92	31,52

Fonte: Autores.

Segundo Watts e Strogatz (1998), uma rede apresenta o efeito *small world* se  $C_{ws}$  é muito maior que  $C_{ra}$  e se o  $\ell$  é comparável com  $\ell_{ra}$ . Nesta definição,  $C_{ra}$  é o coeficiente de aglomeração médio para uma rede aleatória com mesmo grau médio  $z$  e mesmo número de vértices  $n$ . Analogamente,  $\ell_{ra}$  é o caminho mínimo médio para a rede aleatória. De acordo com os resultados da Tabela 3, podemos inferir que todas as redes seguem o modelo *small world*.

Do ponto de vista das redes semânticas de títulos, isto significa que as palavras da rede estão, em termos de comprimento geodésico, muito próximas umas das outras, ou seja, o caminho geodésico de uma palavra

até outra é curto. Significa também uma alta probabilidade de que duas palavras ligadas a uma outra estejam, elas mesmas, ligadas entre si. Esse fato é expresso pelo coeficiente de aglomeração médio.

Por outro lado, para caracterizar melhor as redes, investigamos também os pontos de corte para cada uma delas. Pontos de corte em uma rede são definidos como aqueles vértices cuja retirada da rede faz com que o número de componentes passe a ser maior que 1. A Tabela 2 mostra a quantidade de pontos de corte para as redes.

Observa-se pela Figura 5 que o número de componentes da rede inicial é sempre menor que o número de pontos de corte. Esse resultado é importante na dinâmica da rede, pois mostra uma tendência de a rede se tornar conectada à medida que aumenta o número de títulos nos quais há a ocorrência de palavras iguais. Outro índice a destacar é relação entre o número de pontos de corte ( $P_c$ ) e o número total de vértices da rede ( $n$ ), que designaremos por  $R_r = P_c / n$ . Este índice mede a proporção de vértices que são pontos de corte, escolhidos ao acaso, em relação ao número total de vértices. Como a rede é formada pela ligação entre os títulos (i.e. cliques), a situação mais desfavorável ocorre quando cada título é ligado a outro título por apenas uma palavra (i.e. vértice), como no exemplo apresentado na Figura 1. Nesta topologia, o número máximo de pontos de corte ( $N_T$ ) é igual ao número de títulos ( $T$ ) menos 1 ( $N_T = T - 1$ ). Para que  $R_r$  varie entre 0 e 1, o que possibilita comparar redes de diferentes tamanhos, pode-se normalizar por  $N_T / n$ , obtendo-se simplesmente  $nR_r = P_c / N_T$ . A Figura 6 mostra os resultados desse índice para as redes. Observa-se que a rede da RPM apresenta um índice alto em relação às outras redes. Significa que a rede da RPM é muito susceptível à retirada de palavras em seus títulos, ou seja, as ligações entre os títulos são relativamente fracas. Quando a rede não possui pontos de corte ( $R_r = nR_r = 0$ ), a comparação deve ser feita por outros índices.

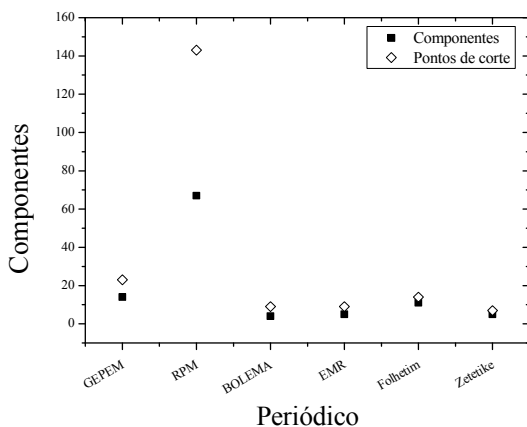


Figura 5 – Relação entre o número de componentes da rede e o número de pontos de corte para a rede conectada

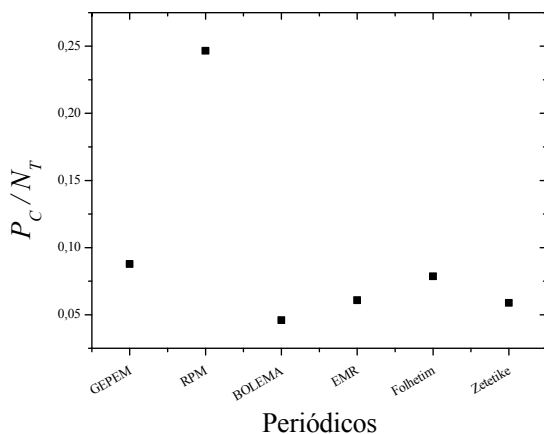


Figura 6 – Relação entre o número de pontos de corte  $P_C$  e o número títulos  $N_T$

Uma importante característica das redes complexas é a sua distribuição de graus, que revela a topologia da rede. Uma rede cuja distribuição de graus se aproxima de uma lei de potência é conhecida em grande número de publicações como *rede livre de escala*. Como se pode observar na Figura 7, as redes estudadas seguem uma lei de potência da forma  $P(k) \sim k^{-\gamma}$ . A Figura 7 também mostra os valores de  $\gamma$  encontrados para as redes conectadas. Uma característica importante das redes



com distribuição livre de escala é que elas são mais robustas em relação à remoção aleatória de vértices que as redes aleatórias de Erdős-Rényi, mas são menos robustas a remoção de um vértice específico com grau alto (Newman et al., 2006, p. 425).

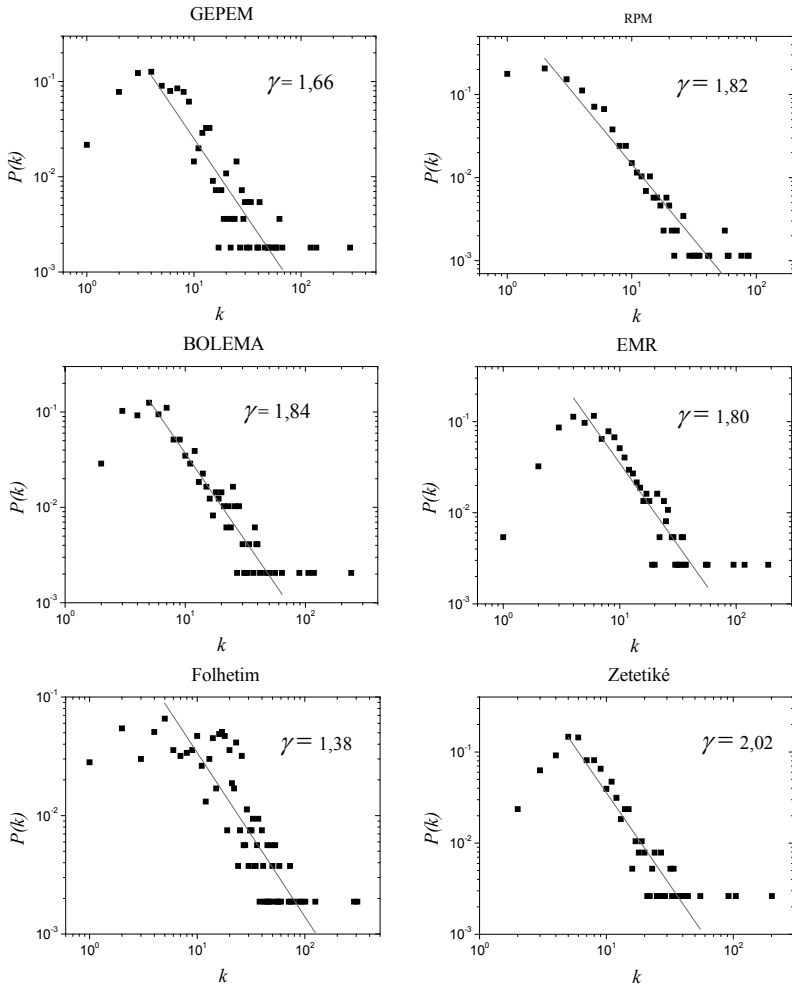


Figura 7 – Distribuição dos graus e valores de  $\gamma$  para as redes conectadas obtidos de  $P(k) \sim k^{-\gamma}$

Redes reais, a exemplo de rede de citações, a Word Wide Web, a Internet e redes metabólicas estudadas por Price (1965) e Barabási e Albert (1999) mostram que a distribuição de graus se aproxima de uma lei de potência. Em uma abordagem de dinâmica de redes, esse comportamento mostra que novos vértices introduzidos na rede tendem a se ligar a vértices com alto grau. Nas redes de títulos, as palavras dos novos títulos associam-se preferencialmente a palavras que já estão ligadas a várias outras. Por exemplo, nas redes estudadas, há uma probabilidade alta de novas palavras ligarem-se à palavra “educação matemática”. Os valores de  $\gamma$  obtidos são menores que aqueles apontados na literatura, que estão entre 2,1 e 4 (Barabási e Albert, 1999), e quando comparados aqueles encontrados por Caldeira (2005) e Teixeira (2007) para redes semânticas. Redes com  $\gamma$  menores que dois recentemente foram estudadas por Seyed-allaei et al. (2006).

Os resultados obtidos dos índices da teoria das redes complexas mostram que as redes semânticas estudadas se caracterizam topologicamente por serem redes *small world* e também livres de escala, uma vez que esses tipos de redes não são excludentes.

### 2.3. Índices da análise de redes sociais

Nesta seção serão apresentados os resultados e discussões com a abordagem da análise de redes sociais. Algumas questões importantes podem ser discutidas com base nessa análise:

- Identificação das palavras que são mais importantes para cada rede de títulos de periódicos;
- Articulação entre as redes;
- Relação com a proposta editorial de cada periódico.

Uma resposta preliminar para a questão das palavras importantes poderia ser: aquela que mais aparece nos títulos. Porém, a simples quantificação não estabelece o mais importante nas análises de redes sociais que é a relação estabelecida entre as palavras. Outras definições e índices serão usados, sobretudo aqueles que expressem a importância das palavras em relação à centralidade.

Na linguagem das redes sociais, vértices são denominados atores, e no caso presente os atores são as palavras significativas de cada título, após os tratamentos realizados. Para padronizar e possibilitar a comparação entre as redes sob a perspectiva da centralidade, as principais palavras até o limite aproximado de 1% da quantidade de palavras (vértices) de cada rede foram

consideradas. Dois tipos de centralidade foram analisados: centralidade de grau e centralidade de proximidade. A Tabela 4 mostra as centralidades analisadas dos principais vértices (i.e. palavras dos títulos) para cada periódico científico e/ou de divulgação matemática na área de educação matemática.

- **Centralidade de grau:** a centralidade de grau foca a importância de um ator nas simples conexões que este estabelece com os atores vizinhos, e é quantificada pelo grau do vértice. Assim, uma palavra na rede é mais importante que outra se estabelece um maior número de vínculos com as palavras vizinhas.

O critério de centralidade de grau mostra que a palavra “matemática” aparece em todas as redes e entre as três maiores frequências, evidenciando sua preferência entre os títulos. Em seguida, aparece “ensino”, com cinco ocorrências, seguida de “professor”, com ocorrência preferencial em quatro redes. Segue a palavra “educação matemática” que ocorre em três redes. Nota-se também que nos periódicos *BOLEMA*, *GPEM* e *Zetetiké* estas quatro palavras ocorrem entre as primeiras, pelo critério estabelecido. Tem-se reconhecido, portanto, um padrão que une as três publicações, cujos títulos possuem um núcleo central formado pelas quatro palavras citadas. Se se excetua a palavra “educaçõamatemática”, pode-se incluir também a *EMR* no grupo;

- **Centralidade de proximidade** – o grau de proximidade de um ator mede o quanto o vértice que representa o ator está próximo de todos os demais vértices da rede. Este índice tenta capturar não só a importância local de uma palavra em relação aos vizinhos mais próximos, mas em relação a toda rede de palavras.

Segundo o critério de centralidade de proximidade, a palavra “matemática” continua a aparecer em todas as redes, seguida de “ensino”, aparecendo em cinco redes, na faixa selecionada. A palavra “professor” ocorre em quatro redes. O resultado é o mesmo daquele da centralidade de grau, eventualmente com alterações no *ranking* das palavras. Para a centralidade de proximidade, porém a palavra “educaçõamatemática” dá lugar a outras como “formação” e “número”; a palavra “formação” que ocorre em duas redes cada. Nota-se agora que o núcleo comum de palavras composto de “matemática”, “ensino”, “professor” aparece nos periódicos *GPEM*, *BOLEMA*, *EMR* e *Zetetiké*.

Tabela 4 – Centralidade de grau e centralidade de proximidade para os periódicos científicos e/ou de divulgação na área de educação matemática

Vértices (i.e. palavras)	Centralidade de grau (%)					Centralidade de proximidade (%)						
	GEPEM	RPM	BOLEMA	EMR	Folhetim	Zetetiké	GEPEM	RPM	BOLEMA	EMR	Folhetim	Zetetiké
matemático *	51,805	10,440	49,489	50,943	53,659	52,880	66,626	39,989	65,200	65,088	66,750	67,434
educaçãomatemática	11,371		24,131			14,398			53,268			
ensino	25,090	7,692	21,677	31,806		27,487	52,868	41,120	52,244	56,127		54,429
professor	22,202		18,200	25,606		24,084	53,121		51,474	55,126		54,978
formação			13,088	15,364					48,998	49,665		
relato	12,093						48,381					
educação							48,254					
ser		8,104			57,598			39,532			68,645	
não					23,452						54,010	
saber					18,949						51,650	
número		11,951			17,656			42,891			51,650	
problema		11,676						42,169				
geométrico		8,242						41,260				
geometria		7,692						38,835				
histórico												50,800

\* Os programas usados para o tratamento dos textos apresentam os nomes sempre no masculino, com exceção daqueles que não fazem parte do dicionário do programa.

Fonte: Autores.

Pode-se parcialmente concluir que as palavras “matemática”, “ensino” e “professor” são aquelas mais importantes, do ponto de vista das duas centralidades, e são comuns a um bloco formado pelos periódicos *GEPEM*, *BOLEMA*, *EMR* e *Zetetiké*. O padrão obtido revela uma convergência na escolha de palavra usadas nos títulos, o que destaca o grupo das outras três publicações. A interpretação da análise de redes sociais aplicada às redes de títulos mostra que as três palavras acima são aquelas que estão mais próximas de suas vizinhas na rede, bem como mais próximas de todas as outras. As demais palavras, principalmente para a centralidade de proximidade, revelam o viés adotado em seus títulos por cada publicação. Por exemplo, enquanto o *BOLEMA* apresenta um viés para “educação matemática”, o *EMR* apresenta para “formação” (Tabela 4).

Ainda sobre centralidades, a observação da Tabela 4 mostra que os índices percentuais da centralidade de grau caem mais rapidamente que aquele para a centralidade de proximidade. O índice que mede este comportamento é a **centralização da rede**. Um índice de centralização em nível de grupo mede a homogeneidade das centralidades dos atores em relação a toda rede. A Figura 8 mostra a centralização de grau  $C_D$  e a centralização de proximidade  $C_C$  para cada rede. Observa-se que os índices para centralização de grau são sempre maiores que os índices para centralização de proximidade. Outro fato é que apenas a *RPM* apresenta centralizações fora da faixa das outras publicações. Destacam-se também as centralizações do *Folhetim* com valores muito próximos.

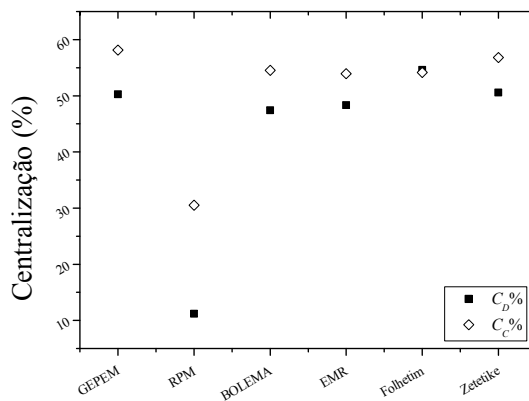


Figura 8 – Centralização de grau e de proximidade para os periódicos

Os índices (exceto da RPM) indicam um vocabulário “padrão” que os autores utilizam para divulgar seus trabalhos. A associação entre o vocabulário usado e a difusão do conhecimento reside no fato de favorecer para o leitor mecanismos de busca por título.

### 3. Considerações finais

A análise dos periódicos em educação matemática feita no presente artigo tem como principal foco a contribuição das redes complexas e da análise de redes sociais no estudo de redes de palavras dos títulos dos artigos publicados em cada periódico. Cabe ressaltar que os títulos não traduzem, necessariamente, o conteúdo tratado e não se pretendeu associar títulos a conteúdos. A contribuição principal é o estudo das redes semânticas de títulos e suas estruturas baseado na análise de redes sociais e no formalismo das redes complexas, de modo a propiciar uma abordagem, qualitativa e quantitativa, da difusão do conhecimento em educação matemática.

Em relação ao componente principal (rede conectada), os índices da abordagem de redes complexas mostram um padrão em todas as redes: enquadram-se no modelo de Watts-Strogatz (*small world*) e exibem um comportamento de distribuição de graus livre de escala. Do ponto de vista das redes complexas, o fato de a rede ser livre de escala, conforme já mencionado, torna a rede robusta à retirada de vértices aleatoriamente, porém, para cada rede individualmente, serão feitas as considerações de vulnerabilidade a partir dos pontos de corte. A seguir são apresentadas, de forma individual, as análises para cada periódico:

1. **GPEM**: apresenta um razoável índice em termos de artigos publicados em cada número de suas edições (5,0). Porém, a existência de 14 componentes na rede indica que as palavras usadas nos títulos não estão muito interligadas, e neste caso ocorrem treze títulos que não se relacionam entre si, nem com os demais do maior componente. Pode-se dizer que há uma forte dispersão nas palavras dos títulos, não apresentando um caráter de uniformidade. A relação entre o número de palavras e o número de títulos (4,8931) é uma das menores. Significa a presença de títulos com poucas palavras, e que não estão associadas a palavras de outros títulos. Quanto à vulnerabilidade da rede, o uso da relação entre o número de pontos de corte e o número total de títulos (8,8%), mantém a coesão da rede com um baixo grau de vulnerabilidade. Os resultados da análise de

redes sociais mostram que o núcleo de palavras importantes dos títulos para a interseção das duas centralidades, compõe-se de “matemático”, “ensino”, “professor” e “relato”. A palavra “educaçãomatemática” compõe o núcleo para a centralidade de grau, mas é substituída por “educação” na centralidade de proximidade. Nesta, “educaçãomatemática” ocupa o 8º lugar. Os índices de centralização ( $C_D = 50,28$  e  $C_C = 58,15$ ) apontam para uma homogeneidade da rede;

**2. RPM:** apresenta o maior índice em termos de artigos publicados em cada número de suas edições (9,5). A presença de muitos componentes na rede (67) indica um critério aleatório para a escolha das palavras dos títulos, pois estas estão muito isoladas, com a presença de 66 títulos que não se relacionam entre si, nem com os demais do maior componente. A relação entre o número de palavras e o número de títulos (3,1034), a menor entre as redes estudadas, mostra uma tendência esperada de que a existência de poucas palavras nos títulos interfira na interligação entre elas na rede. Observando a relação entre o número de pontos de corte e o número total de títulos (24,6%), nota-se que é a maior dentre as redes, o que em termos relativos indica um alto grau de vulnerabilidade. Os resultados da análise de redes sociais mostram que o núcleo de palavras importantes dos títulos para a interseção das duas centralidades compõe-se de “número”, “problema”, “matemático”, “geométrico”, “ser”, “ensino” e “geometria”. Curioso notar que a palavra “educaçãomatemática” não aparece na rede. Quanto à homogeneidade, os índices ( $C_D = 11,20\%$  e  $C_C = 30,53\%$ ) apontam em sentido contrário, ou seja, para uma rede heterogênea em relação às centralidades;

**3. BOLEMA:** situa-se na faixa intermediária em termos de artigos publicados em cada número de suas edições (6,8). A presença de poucos componentes na rede (4) indica que as palavras dos títulos estão muito interligadas, com apenas três títulos que não se relacionam entre si, nem com os demais do maior componente. As palavras dos títulos mantêm certa uniformidade. Embora com uma relação entre o número de palavras e o número de títulos (5,6581) superior àqueles de três outras redes, esta uniformidade traduz-se pelo menor número de componentes. A partir da relação entre o número de pontos de corte e o número total de títulos (4,6%), nota-se que é a menor dentre as redes, o que confirma um baixo grau de vulnerabilidade à retirada aleatória de palavras (vértices). Os resultados da análise de redes sociais mostram que o núcleo de palavras

importantes dos títulos, para a interseção das duas centralidades, compõe-se de “matemática”, “educaçãomatemática”, “ensino”, “professor” e “formação”, nessa ordem, e exhibe uma homogeneidade da rede, quantificada pelos índices de centralização ( $C_D = 47,41\%$  e  $C_C = 54,54\%$ );

4. **EMR**: segue a mesma tendência em termos de artigos publicados em cada número de suas edições (6,7). A presença de poucos componentes na rede (5) indica também que as palavras dos títulos estão muito interligadas, mas neste caso são quatro títulos que não se relacionam entre si, nem com os demais do maior componente. As palavras dos títulos mantêm certa uniformidade. A relação entre o número de palavras e o número de títulos (5,5202) é maior que o de outras duas redes, mas a uniformidade traduz-se pelo número de componentes. O uso da relação entre o número de pontos de corte e o número total de títulos, que não é a menor dentre as redes (6,1%), ainda confirma um baixo grau de vulnerabilidade. Os resultados da análise de redes sociais mostram que o núcleo de palavras importantes dos títulos para a interseção das duas centralidades compõe-se de “matemática”, “ensino”, “professor” e “formação”, nessa ordem, e exhibe uma homogeneidade da rede, quantificada pelos índices de centralização ( $C_D = 48,35\%$  e  $C_C = 53,93\%$ ). É interessante notar que a palavra “educaçãomatemática”, que dá o título à publicação, ocupa os lugares de número 11 e 19 em relação à centralidade de grau e proximidade, respectivamente;

5. **Folbetim**: apresenta um baixo índice em termos de artigos publicados em cada número de suas edições (3,9). A presença de 11 componentes na rede já indica que as palavras usadas nos títulos não estão muito interligadas, e neste caso ocorrem dez títulos que não se relacionam entre si, nem com os demais do maior componente. Há certa dispersão nas palavras dos títulos, não apresentando um caráter de uniformidade. A relação entre o número de palavras e o número de títulos (6,6179) é maior que das outras redes. Significa a presença de títulos com muitas palavras, mas que não estão associadas a palavras de outros títulos. O uso da relação entre o número de pontos de corte e o número total de títulos (7,9%), mostra ainda um grau baixo de vulnerabilidade. Os resultados da análise de redes sociais mostram que o núcleo de palavras importantes dos títulos para a interseção das duas centralidades, compõe-se de “ser”, “matemático”, “não”, “saber” e “número”. Embora algumas palavras sejam não habituais, como “ser”, “não”, “saber”, a rede exhibe uma



homogeneidade, quantificada pelos índices de centralização ( $C_D = 54,61$  e  $C_C = 54,15$ ) que são bem próximos. É interessante notar que a palavra “educaçãomatemática”, que dá o título à publicação, tem centralidade de grau e centralidade de proximidade muito baixos;

**6. Zetetiké:** apresenta o segundo mais baixo índice em termos de artigos publicados em cada número de suas edições (4,6). A existência de 5 componentes na rede indica que as palavras usadas nos títulos são muito interligadas, e neste caso ocorrem quatro títulos que não se relacionam entre si, nem com os demais do maior componente. Pode-se dizer que há uma forte ligação entre as palavras dos títulos, apresentando um caráter de uniformidade. A relação entre o número de palavras e o número de títulos (6,2605) é uma das maiores. Significa que a presença de títulos com muitas palavras contribui para associação entre os títulos. Quanto à vulnerabilidade da rede, o uso da relação entre o número de pontos de corte e o número total de títulos (5,8%), mantém a coesão da rede, com um baixo grau de vulnerabilidade. Os resultados da análise de redes sociais mostram que o núcleo de palavras importantes dos títulos para a interseção das duas centralidades, compõe-se de “matemática”, “ensino”, “professor” e “educaçãomatemática”, acompanhando o padrão de outras redes. Os índices de centralização ( $C_D = 50,59$  e  $C_C = 56,8$ ) apontam para uma homogeneidade da rede.

No geral, a pesquisa mostrou que, inicialmente, podemos agrupar as redes semânticas de palavras usadas nos títulos dos periódicos, em dois grupos: um grupo formado pelos periódicos *GPEM*, *BOLEMA*, *EMR*, *Folhetim* e *Zetetiké* e outro formado pela *RPM*. No primeiro grupo percebe-se uma compatibilidade, tanto nos quantitativos gerais, nos parâmetros das redes complexas e nos índices da análise de redes sociais. Porém, uma análise qualitativa dos resultados de redes sociais aponta ainda para uma classificação em dois grupos, porém um formado pelos periódicos *GPEM*, *BOLEMA*, *EMR*, e *Zetetiké* e outro pela *RPM* e *Folhetim*, levando em conta o caráter diversificado das palavras quanto às centralizações.

## Referências

- BARABÁSI, A. L. e ALBERT, R. (1999). Emergence of Scaling in Random Networks. *Science*, n. 286, pp. 509-512.
- BORGATTI, S. P.; EVERETT, M. G. e FREEMAN, L. C. (2002). *Ucinet 6 for Windows: Software for Social Network Analysis*.
- CALDEIRA, S. M. G. (2005). *Caracterização da Rede de Signos Lingüísticos: Um modelo baseado no aparelho psíquico de Freud*. Dissertação de mestrado em Modelagem Computacional, Centro de Pós-graduação e Pesquisa da Fundação Visconde de Cairu. Salvador, Fundação Visconde de Cairu.
- CALDEIRA, S. M. G.; PETIT LOBÃO, T. C.; ANDRADE, R. F. S.; NEME, A. e MIRANDA, J. G. V. (2006). The network of concepts in written texts. *The European Physical Journal B*, v. 49, pp. 523-529.
- FALOUTSOS, M.; FALOUTSOS, P. e FALOUTSOS, C. (1999). On power-law relationships of the Internet topology. *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pp. 251-262.
- FERRER, I. CANCHO, R. e SOLÉ, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London*, v. 268, pp. 2261-2265.
- JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z. N. e BARABÁSI, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, v. 407, n. 6804, pp. 651-654.
- LILJEROS, F.; EDLING, C. R.; AMARAL, L. A. N.; STANLEY, H. E. e ÅBERG, Y. (2001). The web of human sexual contacts. *Nature*, v. 411, pp. 907-908.
- NEWMAN, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, v. 64, pp. 016131.1-016131.8.
- \_\_\_\_\_. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, v. 64, pp. 016132.1-016132.7.
- NEWMAN, M. E. J.; BARABÁSI, A. L. e WATTS, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton, Princeton University Press.

- PEREIRA, H. B. B.; FREITAS, M. C. e SAMPAIO, R. R. (2007). Fluxos de informações e conhecimentos para inovações no arranjo produtivo local de confecções em Salvador/BA. *DataGramZero*, v. 8, n. 4.
- PRICE, D. J. (1965). Networks of scientific papers. *Science*, v. 149, n. 3683, pp. 510-515.
- SEYED-ALLAEI, H.; BIANCONI, G. e MARSILI, M. (2006). Scale-free networks with an exponent less than two. *Physical Review E*, v. 73, n. 046113, pp. 046113-1-046113-5.
- TEIXEIRA, G. M. (2007). *Redes Semânticas em Discurso Oral: uma Proposta Metodológica baseada na Psicologia Cognitiva utilizando Redes Complexas*. Dissertação de mestrado em Modelagem Computacional, Centro de Pós-graduação e Pesquisa da Fundação Visconde de Cairu. Salvador, Fundação Visconde de Cairu.
- VIEIRA, I. T. (2005). *Small World Networks Models of the Dynamics os HIV Infection*. Doctoral Thesis, Faculty of Engineering, Science & Mathematics, University of Southampton.
- WATTS, D. J. e STROGATZ, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 4, pp. 440-442.

*Recebido em fev./2009; aprovado em mar./2009*