

Redes semânticas em Ciências dos Materiais: uma aplicação da Teoria de Redes

Nascimento, J. O. do.^{1*}; Pereira, H.B.B.^{1,2}; Moret, M. A.^{1,2}; Barbosa, L. P.³; Takiishi, H.³

¹ Departamento de Modelagem Computacional, Centro Universitário SENAI CIMATEC, Salvador, BA, Brasil.

² Universidade do Estado da Bahia, Salvador, BA, Brasil.

³ Instituto de Pesquisas Energéticas e Nuclear - IPEN/USP, São Paulo, Brasil.

* e-mail: jeffersonascimento@gmail.com

Resumo

Uma possibilidade de estudos em sistemas complexos corresponde a utilização da Teoria de redes sociais e complexas. Este artigo tem como objetivo descrever e analisar as redes semânticas baseadas nas palavras-chave dos artigos publicados nos anais do Congresso Brasileiro de Engenharia e Ciências dos Materiais (CBECiMat) referente ao ano de 2010. Alicerçamo-nos nas métricas de estatística básica pertencente às redes complexas, ao índice de centralidade de grau pertencente as redes sociais e apresentamos o método de construção das redes semânticas. As análises indicaram quais as palavras-chave se destacaram (por meio do número de suas conexões) e as topologias verificadas, sugerem redes livre de escala e mundo pequeno. Estes resultados auxiliam na escolha e observância de palavras-chave em futuros trabalhos nas áreas de ciências dos materiais e do evento (materiais cerâmicos, materiais compósitos, materiais metálicos, materiais poliméricos e ensino de materiais), bem como arguir sobre estratégias em mecanismos de busca de trabalhos científicos, tendo como descritores as palavras-chave.

Palavras-chave: Teoria de redes, Redes semânticas, Redes complexas, Ciência dos materiais, CBECiMat.

Abstract

One possibility of studies in complex systems corresponds to the use of Theory of social and complex networks. This article aims to describe and analyze the semantic networks based on the keywords of the articles published in the Brazilian Congress of Engineering and Materials Science (CBECiMat) for the year 2010. We base of on the basic statistical metrics of the networks Complex, the index of centrality of degree belonging to social networks and we present the method of construction of semantic networks. The analyzes indicated which keywords were highlighted (through the number of their connections) and the verified topologies, suggest networks of scale free. These results help in choosing and observing keywords in future works in the areas of the event (ceramic materials, composite materials, metallic materials, polymer materials and materials teaching), as well as to argue about strategies in search engines of scientific works, having its based on keywords.

Keywords: Network Theory, Semantic Network, Complex Network, Materials Science, CBECiMat.

1. Introdução

Para a produção de artigos, dissertações e teses que vislumbram o conteúdo técnico de uma determinada área, a escolha de títulos e palavras-chave pelos autores indicam uma representação coerente com as ideias que serão apresentadas no corpo da referida produção [1]. Desta forma, a presente pesquisa tem como objetivo propor por meio da Teoria de Redes, a aplicação de uma metodologia para descrição e a análise das redes semânticas construídas com base nas palavras-chave dos artigos publicados nos anais

do Congresso Brasileiro de Engenharia e Ciências dos Materiais (CBECiMat) referente ao ano de 2010. A escolha pelo CBECiMat ocorreu em função da sua alcance e impacto científico e, o elevado número de artigos publicados. Para o presente artigo, alicerçamo-nos na Teoria dos Grafos e na supracitada Teoria de Redes, a fim de verificarmos um sistema complexo por meio das conexões entre as palavras-chave das publicações do evento. Assim, construímos as redes semânticas de palavras-chave para cada área de pesquisa do CBECiMat. Realizamos os cálculos dos índices das redes complexas, da centralidade de grau

pertence às redes sociais e a caracterização topológica de cada rede.

2. Materiais e Métodos

A base de dados utilizada correspondeu aos anais do 19º Congresso Brasileiro de Engenharia e Ciência dos Materiais – CBECiMat, realizado em 2010. Foram utilizados 1772 trabalhos para a mineração de 2370 das palavras-chave, para construir as redes semânticas propostas nesta pesquisa. Para representarmos uma rede semântica, utilizamos um grafo $G = (V, E)$ que consiste em uma estrutura matemática composta por dois conjuntos: V (finito e não vazio) e E (relações binárias sobre V) [2]. Os elementos de V são denominados de vértices e os elementos de E são as arestas [2]. As propriedades observadas na análise das redes semânticas propostas neste artigo correspondem àquelas relacionadas aos índices de estatística básica da teoria de redes, caracterizando também as suas topologias. As propriedades são: número de vértices (n), número de arestas (m), grau médio ($\langle k \rangle$), densidade (Δ), coeficiente de aglomeração de um vértice (C_V), coeficiente de aglomeração Médio (C_{ws}), caminho mínimo médio (L) e diâmetro (D), especificados da seguinte maneira:

Número de vértice (n):

$$n = |V| \quad (1)$$

Número de arestas (m):

$$m = |E| \quad (2)$$

Grado médio ($\langle k \rangle$):

$$\langle k \rangle = \frac{1}{n} \sum_i^n k_i \quad (3)$$

onde k_i é o valor da quantidade de vizinhos do vértice i (a quantidade dos vértices adjacentes ou a quantidade de arestas incidentes).

Densidade (Δ):

$$\Delta = \frac{m}{n(n-1)/2} \quad (4)$$

O coeficiente de aglomeração de um vértice V , denominado de C_V :

$$C_V = \frac{2E_v}{k_v(k_v-1)} \quad (5)$$

Coeficiente de aglomeração Médio (C_{ws}):

$$C_{ws} = \frac{1}{N} \sum_{v=1}^N C_v \quad (6)$$

Caminho mínimo médio ou distância geodésica (L):

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \quad (7)$$

onde o termo d_{ij} corresponde a distância geodésica entre os vértices i e j , em termos do número de arestas existentes.

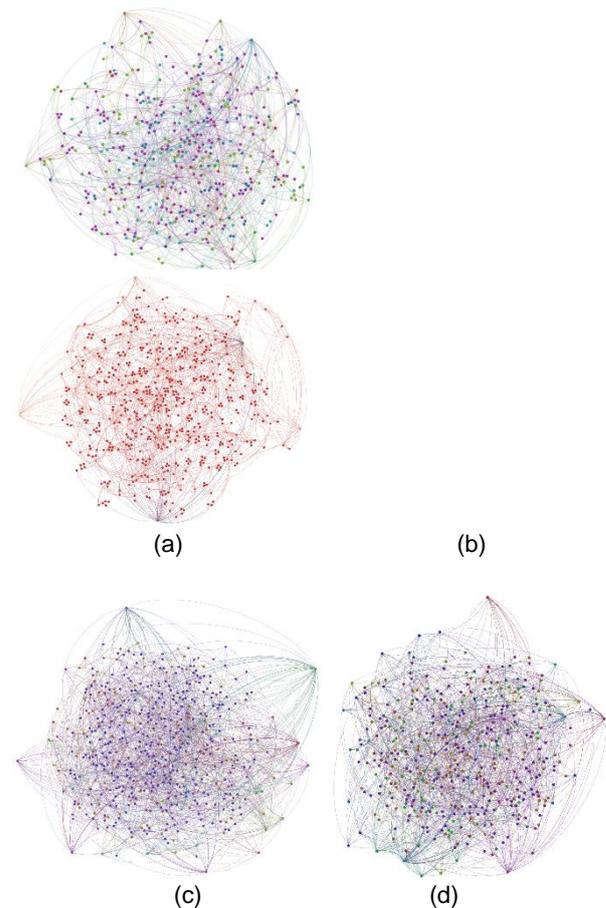
Diâmetro (D):

$$D = \max(d_{ij}) \quad (8)$$

Para que as redes semânticas pudessem ser construídas, realizamos inicialmente um processamento manual nas palavras-chave, com os procedimentos e regras estabelecidas em Pereira et al. [3] e adaptadas conforme Nascimento et al. [1].

3. Resultados e Discussão

Por meio da metodologia proposta de construção de redes semânticas apresentada na seção anterior, apresentamos na Figura 1, as redes baseadas nas palavras-chave dos trabalhos publicados no CBECiMat, referentes às suas cinco áreas:



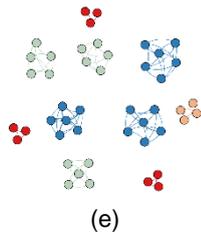


Figura 1 – Redes semânticas baseadas nas palavras-chave dos trabalhos publicados nas áreas de Materiais Cerâmicos (a), Materiais Compósitos (b), Materiais Metálicos (c), Materiais Poliméricos (d) e Ensino de Materiais (e). Fonte: Dos autores.

A rede semântica da Figura 1e - área de ensino de materiais - apresenta a rede completa. É válido destacar que esta rede apresentou dez cliques desconectadas. Não houve nenhuma palavra-chave (vértice) em comum dentre os trabalhos nessa área de pesquisa, diferentemente das demais redes da Figura 1. Os índices de redes complexas para as redes semânticas baseadas em palavras-chave, em cada área pesquisa, estão expostas nas tabelas a seguir:

Tabela 1 – Índices verificados à rede semântica (maior componente) de Materiais cerâmicos e a sua rede aleatória equivalente. Fonte: Autores.

Rede	Propriedade	Valor	Propriedade	Valor
Palavras-chave	$n = V $	441	Maior Componente (%)	57,57
	$m = E $	94	$\langle k \rangle$	4,508
	Δ	0,005	C_{ws}	0,845
	L	5,324	D	11
Rede aleatória eq.	$n = V $	441	Maior Componente (%)	100
	$m = E $	1014	$\langle k \rangle$	4,508
	Δ	0,010	C_{ws}	0,012
	L	4,141	D	9

Tabela 2 – Índices verificados à rede semântica (maior componente) de Materiais Compósitos e a sua rede aleatória equivalente. Fonte: Autores.

Rede	Propriedade	Valor	Propriedade	Valor
Palavras-chave	$n = V $	326	Maior Componente (%)	56,4
	$m = E $	773	$\langle k \rangle$	4,742
	Δ	0,015	C_{ws}	0,851
	L	3,876	D	9
Rede aleatória eq.	$n = V $	326	Maior Componente (%)	100
	$m = E $	776	$\langle k \rangle$	
	Δ	0,014	C_{ws}	0,009
	L	3,846	D	8

Tabela 3 – Índices verificados à rede semântica (maior componente) de Materiais Metálicos e a sua rede aleatória equivalente. Fonte: Autores.

Rede	Propriedade	Valor	Propriedade	Valor
------	-------------	-------	-------------	-------

Palavras-chave	$n = V $	500	Maior Componente (%)	60,31
	$m = E $	1109	$\langle k \rangle$	4,436
	Δ	0,009	C_{ws}	0,837
	L	4,769	D	13
Rede aleatória eq.	$n = V $		Maior Componente (%)	100
	$m = E $	1114	$\langle k \rangle$	4,436
	Δ	0,008	C_{ws}	0,007
	L	4,309	D	9

Tabela 4 – Índices verificados à rede semântica (maior componente) de Materiais Poliméricos e a sua rede aleatória equivalente. Fonte: Autores.

Rede	Propriedade	Valor	Propriedade	Valor
Palavras-chave	$n = V $	413	Maior Componente (%)	70,12
	$m = E $	1028	$\langle k \rangle$	4,978
	Δ	0,012	C_{ws}	0,826
	L	4,205	D	11
Rede aleatória eq.	$n = V $	413	Maior Componente (%)	100
	$m = E $	1039	$\langle k \rangle$	4,978
	Δ	0,012	C_{ws}	0,010
	L	3,905	D	8

Tabela 5 – Índices verificados à rede semântica de Ensino de Materiais. Fonte: Autores.

Rede	Propriedade	Valor	Propriedade	Valor
Palavras-chave	$n = V $	39	Quantidade de Componentes/ Maior Componente (%)	10Comp. / 13,04%
	$m = E $	62	$\langle k \rangle$	4,978
	Δ	0,012	C_{ws}	0,826
	L	4,205	D	11

Apresentamos nos Quadros 2 e Quadro 3, as palavras-chave que se destacaram em termos de centralidade de grau, ou seja, aquelas com maiores importâncias devido às suas conexões com outras palavras-chave. Destacamos nos quadros abaixo a ordem de prioridade das palavras, nome do vértice (palavra-chave) e o grau (k):

Quadro 2 – As dez palavras-chave mais importantes (centralidade de grau) por área do CBECiMat. Fonte: Dos autores.

Materiais Cerâmicos		Materiais Compósitos		Materiais Metálicos	
Vértice	k	Vértice	k	Vértice	k
characterization	29	Composite	66	corrosion	54
recycling	26	Composites	62	microstructure	46
Alumina	24	Mechanical	30	steel	40

		properties			
synthesis	20	Polypropylene	27	welding	32
Cathode	20	Epoxy	23	heattreatment	25
Waste	19	Carbon nanotubes	21	titanium	24
Clay	18	Nano composites	21	fatigar	24
nanoparticles	18	Recycling	19	mechanical properties	23
thin films	16	Cellulose	15	aisi	23
Sintering	16	Nano composite	14	aluminum	22

Quadro 3 – As dez palavras-chave mais importantes (centralidade de grau) por área do CBECiMat. Fonte: Dos autores.

Materiais Poliméricos		Ensino de Materiais	
Vértices	k	Vértices	k
nanocomposites	43	lighting	5
poly	43	artworksreservation	5
polypropylene	40	led	5
recycling	40	luminance	5
nanocomposite	28	color	5
blends	27	energyefficiency	5
polyamide	21	aneroid	4
ftir	21	vacuum	4
polymers	20	dissect	4
aging	20	automation	4

As informações nos quadros anteriores retratam, a título de exemplo, que para a rede semântica de matérias cerâmicas, o vértice com maior centralidade de grau foi “characterization”. As palavras “Recycling”, “Alumina” e “synthesis”, seguem como segundo, terceiro e quarto vértices, respectivamente, mais importantes na rede, até a décima palavra-chave mais importante, que correspondeu a “sintering”. Tal observação pode ser realizada para as demais áreas verificadas nos Quadros 1 e Quadro 2. Destacamos que para a área de ensino de materiais, as dez palavras-chave indicadas, refletem apenas a importância em duas cliques distintas e sem conexões entre elas. A Figura 2 apresenta a distribuição de graus realizada à rede semântica de Materiais Cerâmicos. As redes estudadas sugerem um padrão de uma lei de potência, da forma $P(k) \sim k^{-\gamma}$, semelhantes ao da Figura 2, com exceção à área de Ensino de Materiais:

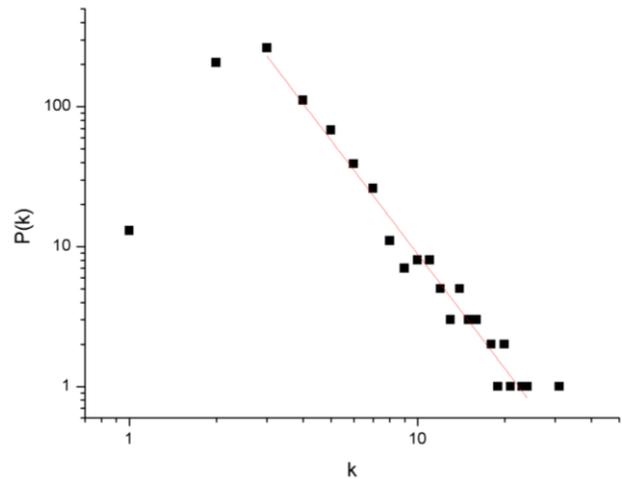


Figura 2: Distribuição de graus da rede semântica de Materiais Cerâmicos (com $\gamma = 2,70646$ e o ajuste $R^2 = 0,9744$).

Pelas considerações já expostas, não realizamos a distribuição de graus da área de ensino de materiais. De acordo com os gráficos da Figura 2, há indícios de que as redes apresentam um comportamento de redes livres de escala. Desta forma, os trabalhos investigados nas áreas de discussão do CBECiMat, apresentam uma adesão preferencial por certos vértices na rede, com a presença de *hubs* (poucos vértices conectando muitos outros) [4]. Assim, novas publicações terão uma tendência em utilizarem as palavras-chave listadas nos Quadros 1 e Quadro 2. O coeficiente angular da distribuição de graus da área de materiais poliméricos, apresentou um valor $\gamma = 1,87989$, menor do que o sugerido por Barabási e Albert [4]: $2,1 \leq \gamma \leq 4$. Contudo, Seyed-allaei et al. [5] e Fadigas et al. [6] também encontraram $\gamma < 2$, para redes *Scale Free* (livres de escala). Conforme o método sugerido por Watts e Strogatz [7], realizamos uma comparação entre a rede semântica de palavras-chave (maior componente) e a sua rede aleatória equivalente (rede com o mesmo número de vértice e grau médio). Nas Tabelas de 1 a 4, apresentamos os valores das propriedades das redes estudadas. Então, diante dos resultados apresentados, verificamos que a rede semântica, apresenta o fenômeno *Small-World*.

4. Considerações finais

Por meio das análises das redes semânticas baseadas nas palavras-chave pertencentes aos trabalhos publicados no CBECiMat de 2010, com os índices da teoria de redes e a caracterização de suas topologias, as redes indicaram características de serem livre de escala. Esta conclusão implica no fato da ocorrência de crescimento e a da adesão preferencial das redes semânticas por algumas palavras-chave específicas, como as elencadas nos Quadro 1 e Quadro 2. Novas publicações poderão surgir aderindo àquelas palavras-chave com maiores centralidades de grau. As redes

livres de escala são resistentes à retirada de vértices de forma aleatória e vulneráveis à retirada de vértices com alta medida de centralidade de grau. Então, caso retirássemos das redes semânticas as palavras-chave elencadas nos quadros supracitados, verificaríamos um fenômeno semelhante ao da Figura 1e: as redes se dividiriam em um grande número de componentes. Assim, ocorreriam dificuldades em se realizarem buscas das publicações tendo como descritores as palavras-chave, devido à falta de conexão dentre as cliques.

5. Agradecimentos

Jefferson Nascimento agradece à FAPESB pelo suporte financeiro parcial devido à bolsa de doutorado (BOL170/2015). Marcelo A. Moret e Hidetoshi Takiishi agradecem ao CNPq pelo suporte financeiro parcial oriundo de sua bolsa de Produtividade em Pesquisa, respectivamente, (No. 304454/2014-1) e (303811/2015-3).

6. Referências

- [1] NASCIMENTO, J. O. do; PEREIRA-GUIZZO, C. S.; MOREIRA, D. M.; MONTEIRO, R. L. S.; PEREIRA, H. B. B.; MORET, M. A.; "Redes Sociais e Complexas: um modelo computacional para a investigação da pós-graduação Brasileira em Ensino de Física", p. 110-114. In: **Anais do VII Encontro Científico de Física Aplicada [=Blucher Physics Proceedings, v.3 n.1]**. São Paulo: Blucher, 2016. ISSN 2358-2359, DOI 10.5151/phypro-vii-efa-027. Disponível em: <<http://www.proceedings.blucher.com.br/article-details/redes-sociais-e-complexas-um-modelo-computacional-para-a-investigao-da-ps-graduao-brasileira-em-ensino-de-fsica-25262>>. Acesso em 18 jul. 2017.
- [2] GROSS, J. L, YELLEN, J. Graph theory and its applications. CRC press, Boca Raton FL USA 2005.
- [3] PEREIRA, H. B. B., FADIGAS, I. S., SENNA, V., MORET, M. A. "Semantic networks based on titles of scientific papers". *Physica A* 390, p. 1192-1197, 2011.
- [4] BARABÁSI, A. L. E ALBERT, R.. "Emergence of scaling in random networks". *Science*, n. 286, pp. 509-512, 1999.
- [5] H. SEYED-ALLAEI, G. BIANCONI & M. MARSILI. Scale-free networks with an exponent less than two. *Phys. Rev. E* 73, 046113, 2006.
- [6] FADIGAS, I. S., CASAS, T. H. P., SENNA, V., MORET, M. A, PEREIRA, H. B. B. "Análise de redes semânticas baseada em títulos de artigos de periódicos científicos: o caso dos periódicos de divulgação em educação matemática". *Educação Matemática Pesquisa*, 11, p. 167-193, 2009.
- [7] WATTS, Duncan J.; STROGATZ, Steven H. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 6684, p. 440-442, 1998.

