

Redes Semânticas Complexas baseadas em abstracts

Nascimento, J. O. do^{1*}; Pereira, H. B. B^{1,2}; Moret, M. A^{1,2}

1 Programa de Modelagem Computacional, CIMATEC, Salvador, BA, Brasil.

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

* e-mail: jeffersonascimento@gmail.com

Resumo

Ao ser elaborada uma obra científica, faz-se necessário a construção e apresentação de seu resumo (*abstract*), correspondendo a uma condensação sucinta do trabalho escrito, expondo de forma direta as nuances apresentadas no corpo do texto. Como possibilidades de buscas de resultados das interações semânticas em produções científicas, a fim de que possamos estudar as propriedades emergentes resultantes destas interações, podemos iniciar as análises utilizando os seus respectivos resumos (*abstracts*). Diante da diversidade de técnicas científicas para o estudo de possíveis padrões que possam ocorrer nos resumos, utilizamos a Teoria dos Grafos e a Teoria de Redes, por meio de redes semânticas complexas. Este artigo tem como objetivo apresentar o método de construção de um modelo computacional baseado em resumos de obras científicas, Calcular os índices estatísticos das redes complexas e as medidas de centralidades pertence às redes sociais. Verificar a topologia das redes semânticas e as implicações fornecidas pelo modelo computacional para o contexto principal da fonte de dados à área de conhecimento investigada, podendo ser a Física, Ensino de Física, Engenharia, Ciências dos Materiais, etc. Desta forma há uma alta probabilidade de podemos verificar a forma da difusão do conhecimento nas análises de redes baseadas em resumos

Palavras-chave: Redes Semânticas, Redes Complexas, Redes Sociais, Física.

Abstract

When a scientific work is elaborated, it is necessary the construction and presentation of its abstract, corresponding to a brief condensation of the written work, exposing in a direct way the nuances presented in the body of the text. As possibilities of searching the results of semantic interactions in scientific productions, in order to study the emergent properties resulting from these interactions, we can start the analyzes using their respective abstracts. Faced with the diversity of scientific techniques for the study of possible patterns that may occur in abstracts, we use Graph Theory and Network Theory, through complex semantic networks. This article aims to present the method of constructing a computational model based on abstracts of scientific works, calculate the statistical indexes of complex networks and measures of centralities belongs to social networks. To verify the topology of the semantic networks and the implications provided by the computational model for the main context of the data source to the researched area of knowledge, being it Physics, Physics Teaching, Engineering, Materials Science, etc. In this way, there is a high probability that we can verify the form of the diffusion of the knowledge, in the analyzes of networks based on abstracts

Keywords: Semantic Networks, Complex Networks, Social Networks, Physics.

1. Introdução

Como possibilidades de buscar padrões semânticos em produções científicas, podem ser realizados estudos tendo como base as palavras que constituem os seus títulos, as palavras-chave e os resumos (*abstracts*) [1]. Desta forma tais pesquisas

apresentam como alicerce a Teoria de Redes [1]. O presente artigo tem como objetivo propor um método de estudo em obras científicas, por meio da modelagem computacional. Assim investigar redes semânticas baseadas em resumos, como uma forma de verificar um determinado padrão em obras científicas pertencentes a uma determinada área de pesquisa, como a Física, por exemplo. Assim como em

Nascimento et al. [2] e Nascimento, Pereira e Moret (2017) [1], a presente pesquisa apresenta como arcabouço teórico-científico a Teoria dos Grafos e a Teoria de Redes – Redes complexas e redes sociais. Nesse viés, elencamos alguns trabalhos que se destacam nos estudos relacionados às redes semânticas, como a pesquisa de Fadigas et al. [3], Pereira et al. [4], Fadigas e Pereira [5], Nascimento et al. [2], Pereira et al. [6], Nascimento, Pereira e Moret [1] e Nascimento et al. [7].

Este artigo está organizado em 4 seções. Na segunda, apresentaremos os materiais e métodos utilizados para a construção da rede semântica baseada em resumos (abstracts). Na terceira, os resultados da presente pesquisa são apresentados. Na quarta seção, as considerações finais do presente artigo. Por fim, na última sessão, as referências que embasam a presente pesquisa.

2. Materiais e Métodos

Os dados utilizados para as construções das redes semânticas baseadas em resumos são os disponibilizados em Salem e Kawamura [8,9]. Para que as redes semânticas baseadas em resumos pudessem ser construídas e representadas, alicerçamo-nos na Teoria dos Grafos. Um grafo $G = (V, E)$ consiste em uma estrutura matemática composta por dois conjuntos: V (finito e não vazio) e E (relações binárias sobre V) (GROSS; YELLEN, 2005) [10]. Os elementos de V são denominados de vértices e os elementos de E são as arestas. O primeiro conjunto (V) é finito e não vazio e o segundo (E) corresponde as possíveis interações binárias sobre V (GROSS; YELLEN, 2005).

Os índices estatísticos pertencentes à teoria de redes calculados, são os indicados em Pereira et al. (2011) e Nascimento et al. (2016): número de vértices (n), número de arestas (E), grau médio ($\langle k \rangle$), densidade (Δ), coeficiente de aglomeração (C_v), coeficiente de aglomeração médio (C_{ws}), caminho mínimo médio (L) e diâmetro (D). Em relação ao pré-processamento das palavras que constituem os resumos, utilizamos as regras apresentadas em Pereira et al. (2011) [4] e Nascimento, Pereira e Moret (2017) [1].

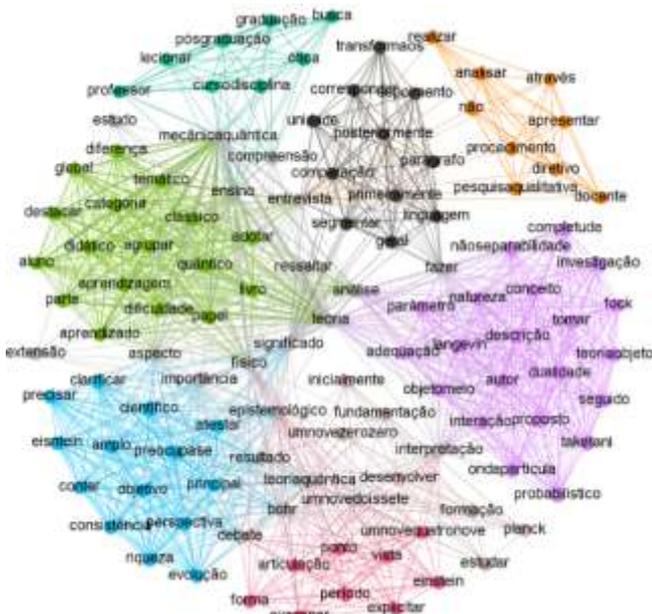
Para a construção da rede semântica foram utilizados os softwares indicados Fadigas et al. [3], Pereira et al. [4], Nascimento et al. [2], Nascimento, Pereira e Moret [1] e Nascimento et al. [7]. A título de exemplo, utilizando os dois resumos de dissertações e teses na área de Física (ensino) presentes em Salem e Kawamura [8,9], para a construção da rede semântica presente na Figura 1.

Resumo 1
Estuda o desenvolvimento epistemológico do conceito de entropia e da 2a. lei da termodinâmica, desde sua origem, com Clausius, até sua importância para o estabelecimento da quantização de energia por Planck. Enfoca nesse estudo a introdução, por Boltzman, da análise probabilística na interpretação do conceito de entropia e as consequentes mudanças na teoria física. Utiliza artigos originais de Clausius, Boltzmann, Zermelo e Planck, assim como artigos de historiadores da ciência como Blackmore, Brush, Hermann, Jammer, Klein, Kuhn e Pais, e de alguns físicos como Bohm, Gibbs, Schenberg e Sommerfeld. Com o intuito de fornecer subsídios tanto a professores de física dos três graus de ensino como a estudantes de terceiro grau, apresenta formas de aproveitamento do trabalho, afirmando: a importância da abordagem histórica e epistemológica na ampliação da compreensão dos conceitos científicos; a necessidade de se introduzir o pensamento da física moderna no segundo grau e a necessidade de se incorporar disciplinas de história e filosofia da ciência no ensino de física em nível superior.

Resumo 2
Busca uma compreensão do ensino de Mecânica Quântica sob a ótica de professores que lecionaram essa disciplina em cursos de graduação e pós-graduação. Através de procedimentos de pesquisa qualitativa, apresenta e analisa entrevistas não diretivas realizadas com cinco docentes. Primeiramente, segmenta as entrevistas em parágrafos que correspondem a "unidades de significado" e transforma-os posteriormente em uma linguagem mais geral para fazer comparações entre os vários depoimentos. Em uma análise mais global, agrupa as entrevistas em três aspectos temáticos - a mecânica quântica e outras teorias físicas, o ensino de mecânica quântica e a aprendizagem de mecânica quântica - e nessas categorias, destaca diferenças no ensino entre as teorias clássica e quântica, o papel dos livros didáticos adotados e as dificuldades de aprendizado por parte dos alunos.

Na Figura 1 é apresentado um exemplo de construção da rede semântica baseada em resumos de dissertações e teses na área de física (ensino):

para o período de 1972-1992 das dissertações e teses na área de física (ensino):



examinar	V
debate	N
bohr	N+Pr
einstein	N+Pr
interpretação	N
teoriaquântica	NOTFOUND
período	N
umnovedoisete	NOTFOUND
umnovetrês	NOTFOUND
forma	N
explicitar	V
articulação	N
ponto	N
vista	N
físico	N
epistemológico	A

inicialmente	ADV
estudar	V
formação	N
teoriaquântica	NOTFOUND
resultado	N
planck	N+Pr
umnovetrês	NOTFOUND
fundamentação	N
desenvolver	V
bohr	N+Pr
umnovedoisete	NOTFOUND

Figura 1: Rede semântica baseada nos Resumo 1 e Resumo 2.

Conforme a utilização dos softwares UNITEX e Ambisin, abaixo da rede semântica na Figura 1, apresentamos duas cliques e seus vértices (palavras) com a sua respectiva identificação de categoria: N para substantivo, A para adjetivo, N+Pr para substantivo próprio, V para verbo, ADV para advérbios, NOTFOUND para palavras não encontradas no dicionário usado pelos softwares). Por meio da construção da rede e das análises dos vértices e métricas pertencentes às Redes Complexas e Sociais, podemos fazer as análises nas obras científicas de interesse em busca de, por exemplo, como se deu o fluxo de informação em uma determinada área; no caso, a física.

3. Resultados

Conforme o método de construção de redes semânticas baseadas em resumos descritos na sessão anterior, apresentamos na Figura 2 a rede construída

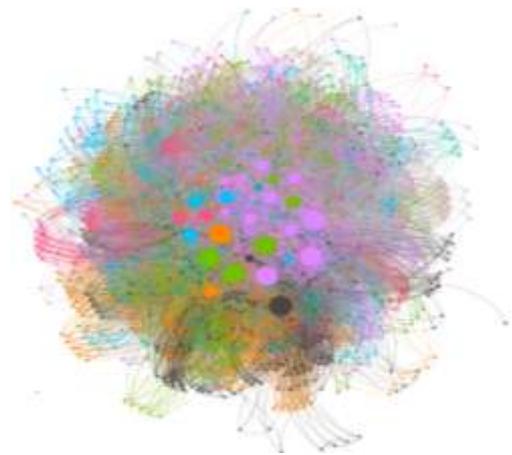


Figura 2: Rede semântica baseada em resumos de dissertações e teses em Física (ensino).

Na tabela 1 apresentamos os índices de redes complexas calculados para a rede semântica baseada em resumos da Figura 2:

Tabela 1: índices pertencentes às redes complexas.

Rede	Propriedade	Valor	Propriedade	Valor
Abstracts (Maior Comp.)	$n = V $	1564	Maior Componente (%)	99,94
	$m = E $	31329	$\langle k \rangle$	40,063
	Δ	0,026	C_{ws}	0,752
	L	2,335	D	4
Rede aleatória equiv.	$n = V $	1564	Maior Componente (%)	100
	$m = E $	31898	$\langle k \rangle$	40,063
	Δ	0,013	C_{ws}	0,026
	L	2,310	D	3

Conforme o método proposto por Watts e Strogatz [11], realizamos uma comparação entre os dados entre a maior componente da rede baseada em resumos com a sua rede aleatória equivalente – uma rede aleatória construída com os mesmos valores de vértices e grau médio da maior componente. Desta forma verificamos que a rede da Figura 2 apresenta o fenômeno *Small-World* (fenômeno mundo pequeno).

Tabela 2: Centralidade de Grau dos vinte vértices (hubs) mais conectados na rede semântica.

Vértices (V)	Grau (k)	Vértices (V)	Grau (k)		
1	Física	586	11	Concluir	328
2	Aluno	489	12	Conhecimento	310
3	Ensino	447	13	Conteúdo	295

4	Ser	438	14	Utilizar	292
5	Apresentar	419	15	Não	286
6	Professor	406	16	Analisar	283
7	Ciência	399	17	Estudo	277
8	Conceito	398	18	Aspecto	276
9	Proposta	350	19	Segundo Grau	272
10	Científico	345	20	Concepção	264

A distribuição de graus da rede semântica da Figura 2 é o apresentado a seguir:

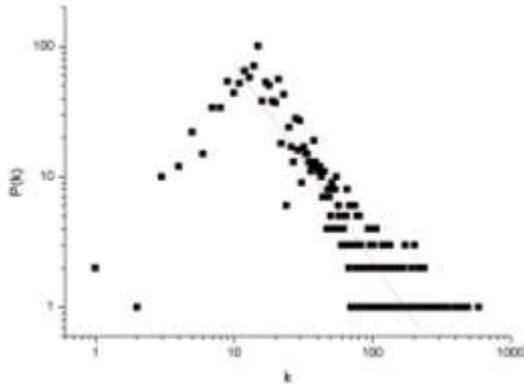


Figura 3: Distribuição de graus da rede semântica baseada em resumos (com $\gamma = 1,5147$ e o ajuste $R^2 = 0,82346$).

Conforme Figura 3, a topologia da rede (Figura 1), sugere seguir uma lei de potência da forma $P(k) \sim k^{-\gamma}$. Uma das características em redes Scale Free é a presença hubs. Os hubs constituem poucos vértices aglomerando muitos outros vértices. Assim, há indícios da rede semântica baseada em resumos, corresponder a uma rede *Scale Free* (rede livre de escala). Este resultando ratifica a presença de *hubs* presente na rede semântica, alguns deles elencados na Tabela 2. O valor do coeficiente angular γ encontrado ser menor do que o verificado por Barabási e Albert [12] ($2,1 \leq \gamma \leq 4$), há na literatura científica redes semânticas com o valor de γ menor do que o indicado pelos supracitados autores. Citamos as obras de Fadigas et al. [3], Seyed-allaei et al. [13], Nascimento et al. [7] e Nascimento, Pereira e Moret [14]. Por fim, ao analisarmos os hubs presentes na tabela 2, verificamos que a rede semântica baseada em resumos de dissertações e teses na área de Física (ensino) no período de 1972 a 1992, sugere que: os trabalhos científicos objetivaram pesquisas em metodologias para o Ensino de Física. O nível de ensino com maior ênfase nas pesquisas em termos de posições de metodologia de ensino, correspondeu a educação básica. Também estiveram presentes nos trabalhos de

mestrado e doutorado, teorias de aprendizagens alicerçadas em Piaget e David Ausubel.

4. Agradecimentos

Jefferson Nascimento agradece à FAPESB por meio do apoio financeiro parcial recebido através da bolsa de doutorado (BOL170/2015). Marcelo A. Moret agradece ao CNPq pelo suporte financeiro parcial oriundo de sua bolsa de Produtividade em Pesquisa (No. 304454/2014-1).

5. Referências

- [1] NASCIMENTO, J. O. do.; PEREIRA, H. B. B.; MORET, M. A.; "Redes semânticas baseadas em palavras-chave do Ensino de Física Brasileiro: uma comparação nos métodos de pré-processamento dos dados", p. 122-127. In: . São Paulo: Blucher, 2017. ISSN 2358-2359, DOI 10.5151/phypro-viii-efa-28.
- [2] 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
- [3] 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.
- [4] 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.
- [5] FADIGAS, I. S.; PEREIRA, H. B. B. A network approach based on cliques. *Physica A: Statistical Mechanics and its Applications*, v. 392, n. 10, p. 2576-2587, 2013.
- [6] PEREIRA, H. B. B. et al. Density: A measure of the diversity of concepts addressed in semantic networks. **Physica A: Statistical Mechanics and its Applications**, v. 441, p. 81-84, 2016.
- [7] NASCIMENTO, J. O. do; MONTEIRO, R. L. S.; MOREIRA, D. M.; MORET, Marcelo A.; PEREIRA, H. B. B. Semantic networks of keywords from Brazilian dissertations on physics teaching. **Discontinuity, Nonlinearity and Complexity**, 2018. (PRELO).
- [8] SALEM, S.; KAWAMURA, M. R. D. Ensino de Física no Brasil: catálogo analítico de dissertações e teses (1972-1992). São Paulo: Instituto de Física da USP / PROFIS, 1992. 243 p.
- [9] SALEM, S.; KAWAMURA, M. R. D. Ensino de Física no Brasil: catálogo analítico de dissertações e teses (1992-1995) São Paulo: s.n., 1996
- [10] GROSS, J. L.; YELLEN, J. Graph theory and its applications. CRC press, Boca Raton FL USA 2005.
- [11] WATTS, Duncan J.; STROGATZ, Steven H. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 6684, p. 440-442, 1998.
- [12] BARABÁSI, A. L. E ALBERT, R.. "Emergence of scaling in random networks". *Science*, n. 286, pp. 509-512, 1999.
- [13] SEYED-ALLAEI, H, BIANCONI & M, G. MARSILI. Scale-free networks with an exponent less than two. *Phys. Rev. E* 73, 046113, 2006.
- [14] NASCIMENTO, J. O. do; PEREIRA, H. B. B; MORET, M. A. Grafos e Teoria de Redes: uma análise do Ensino de Física Brasileiro no período 1972-2006 por meio de cliques de palavras-chave. *REVISTA CEREUS*, v. 10, p. 315-339, 2018. Disponível em: <<http://ojs.unirg.edu.br/index.php/1/article/view/2278/675>>. Acesso em: 07 ago. 2018.

