

Redes semânticas baseadas em palavras-chave do Ensino de Física Brasileiro: uma comparação nos métodos de pré-processamento dos dados

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

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

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

* e-mail: jeffersonascimento@gmail.com

Resumo

As palavras-chave dos trabalhos científicos correspondem a uma importante e inicial demonstração do conteúdo técnico que se discutirá no texto de uma produção científica. Este artigo tem como objetivo analisar, por meio da teoria de redes, as redes semânticas baseadas em palavras-chave dos trabalhos *Stricto Sensu* em Ensino de Física entre 1996-2006. Realizamos uma discussão em relação as regras de pré-formatação publicadas em Pereira et al. (2011), as quais são utilizadas antes do processamento dos dados por meio de *softwares* para a obtenção das redes. Duas redes semânticas baseadas em palavras-chave foram desenvolvidas, observando a supracitada discussão do método de construção, no pré-processamento. Calculamos os índices estatísticos das redes complexas (na forma estática) e, a centralidade de grau pertencente às redes sociais, para ambas as redes. Diante disto, realizamos uma comparação dos resultados, relacionando-os com os supracitados métodos. As redes apresentaram resultados que sugerem uma Lei de Potência, comportamento que indica um modelo Livre de Escala (*Scale Free*) e, também foi observado o fenômeno (*Small-World*). Em ambas as redes, nota-se que a maioria dos assuntos que se destacaram estão relacionados com a formação do professor de física e não com metodologias para o Ensino de Física.

Palavras-chave: Sistemas Complexos, Teoria de Redes, Redes Semânticas, Ensino de Física.

Abstract

The keywords of the scientific papers correspond to an important and initial demonstration of the technical content that will be discussed in the text of a scientific production. This article aims to analyze, through the theory of networks, the semantic networks based on keywords of *Stricto Sensu* works in Physics Teaching between 1996-2006. We conducted a discussion regarding the pre-formatting rules published in Pereira et al. (2011), which are used before data processing through software to obtain the networks. Two semantic networks based on keywords were developed, observing the aforementioned discussion of the construction method, in the pre-processing. We calculate the statistical indices of the complex networks (in static form) and the centrality of degree belonging to social networks, for both networks. Therefore, we compare the results with the above-mentioned methods. The networks presented results that suggest a Power Law, behavior that indicates a Scale Free model, and also the phenomenon was observed (*Small-World*). In both networks, it is noticed that the majority of the subjects that stood out are related to the formation of the teacher of physics and not with methodologies for the Teaching of Physics.

Keywords: Complex Systems, Network Theory, Semantic Networks, Physics Teaching.

1. Introdução

As palavras-chave e os títulos de trabalhos científicos correspondem a uma demonstração de um conteúdo técnico que se discutirá no texto de uma produção científica [1]. Como possibilidades de buscas

de agrupamentos semânticos em produções científicas, podem ser realizados estudos, por meio das palavras que constituem as palavras-chave, tendo como alicerce a Teoria dos Grafos e a de Redes (Redes Sociais e Complexas). A partir da premissa que cada título e/ou conjunto de palavras-chave constituem uma clique [2], tendo como vértices suas respectivas palavras (e as

arestas são as ligações entre elas) podem ser estudadas as possíveis interações entre diferentes trabalhos científicos, seja por época, linhas de pesquisas, preferências temáticas locais (regional) e globais (nacionais), colaboração entre regiões, coautores, etc. Uma clique é um tipo de rede cujo elemento fundamental não é o vértice, senão um conjunto de n vértices que estão mutuamente ligados entre si [3].

Assim o objetivo deste artigo é o analisar, por meio da teoria de redes, as redes semânticas baseadas em palavras-chave dos trabalhos *Stricto Sensu* em Ensino de Física entre 1996-2006, comparando os resultados com as regras de pré-formatação publicadas em Pereira et al. [4] e adaptadas em Nascimento et al. [1]. Desta forma promover uma alteração na pré-formatação realizada em Nascimento et al. [1] e comparar os resultados das redes semânticas resultantes. Apresentaremos os índices de estatística básica pertencentes as Redes Complexas; a centralidade de grau pertencentes as redes sociais e também, a distribuição de graus das redes.

Este trabalho está organizado em 4 seções. Na segunda, apresentaremos os materiais e métodos utilizados para a construção das redes semânticas. Na terceira, os resultados e as discussões. Na quarta seção as considerações finais do presente artigo.

2. Materiais e métodos

O material utilizado como fonte de dados foi o relatório de Salem e Kawamura [5], que apresenta informações pertencentes as dissertações, teses de doutorado e livre docência especificamente na temática de Ensino de Física no período de 1996 à 2006.

Para alcançarmos os objetivos da construção das redes semânticas, descrevemos um grafo $G = (V, E)$, que corresponde a uma estrutura matemática que consistente em dois conjuntos: V (finito e não vazio) e E (relações binárias sobre V) [6]. Os elementos de V são denominados de vértices e os elementos de E , são as arestas. As propriedades observadas das redes complexas, para análise da rede semântica de palavras-chave deste artigo, são: número de vértices (n), número de arestas (E), grau médio ($\langle k \rangle$), densidade (Δ), coeficiente de aglomeração (C_v) e coeficiente de aglomeração médio (C_{ws}) caminho mínimo médio (L) e diâmetro (D), especificados e calculados conforme Pereira et al. [4] e Nascimento et al. [1].

Em relação às métricas de redes sociais, as análises realizadas foram por meio das observações das centralidades de grau. Esse índice indica a importância que um determinado vértice tem, por meio de suas conexões, com vértices vizinhos conectados a ele [2].

Para que a rede pudesse ser construída, realizamos um tratamento manual nas palavras-chave, conforme Quadro 1:

Quadro 1: Regras para a formatação das palavras-chave. Fonte: Adaptado de Pereira et al. [3].

1	Cada conjunto de palavras-chave corresponde a uma sentença;
2	Os sinais gráficos de cada sentença como o ponto e vírgula, ponto de interrogação, ponto de exclamação e reticências são eliminados;
3	Os nomes devem formar uma palavra simples. Por exemplo, "Bose-Einstein" e "Albert Einstein", respectivamente, tornar-se-ão: "boseeinstein" e "alberteinstein".
4	Números ordinais devem ser escritos da seguinte forma, por exemplo, "8ª" e "1º", tornar-se-ão, respectivamente: "oitava" e "primeiro".
5	Os números devem ser escritos de forma textual, por exemplo, "1972" e "1995", tornar-se-ão "umnovesetedeois" e "umnovenovecinco".
6	Palavras compostas devem ser consideradas como uma só palavra, por exemplo, "ensino de ciências", tornar-se-á, "ensinociências".
7	Palavras escritas de formas incorretas, devem ser corrigidas.
8	As palavras que se repetem no conjunto de palavras-chave, devem ser retiradas, permanecendo apenas uma.
9	Linguagem especializada deve ser mantida, sempre que possível.
10	Palavras que apresentam mútua importância, devem se tornar apenas uma, por exemplo, "Física Quântica", tornar-se-á, "FísicaQuântica".
11	Palavras escritas em linguagem diferente do Português, devem ser traduzidas.

Seguindo as recomendações de Nascimento et al. [1] no que se refere aos recursos computacionais foram utilizados os procedimentos em conformidade com Fadigas et al. [2], Pereira et al. [4] e Caldeira [7]. Cabe ressaltar que as ambiguidades são eliminadas, seguindo o algoritmo desenvolvido em Caldeira [7].

3. Resultados e Discussão

Por meio da metodologia referenciada no tópico anterior, Nascimento et al. [1] apresenta a seguinte rede semântica baseada em palavras-chave de dissertações e teses em Ensino de Física no Brasil, conforme Figura 1:

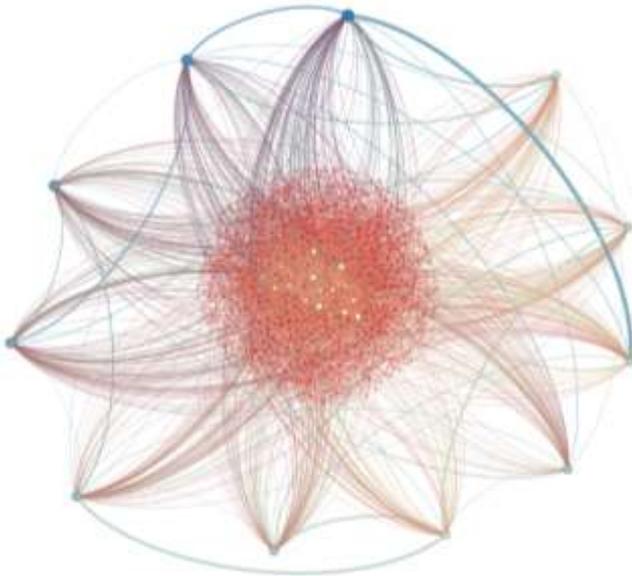


Figura 1: Rede semântica baseada em palavras-chave dos trabalhos em Ensino de Física. Fonte: Nascimento et al. [1].

De acordo com Nascimento et al. [1] as propriedades das redes complexas obtidas em relação a rede semântica da Figura 1, estão contidas na tabela 1:

Tabela 1: Propriedades pertencentes à Rede Semântica. Fonte: Nascimento et al. [1].

Propriedade	Valor	Propriedade	Valor
$n = V $	669	Maior Componente (%)	98,95
$m = E $	5910	$\langle k \rangle$	8,8
Δ	0,013	C_{ws}	0,74
L	3,06	D	6

Utilizando o método apresentado por Watts e Strogatz [8], foi realizada a comparação entre a maior componente da rede semântica baseada em palavras-chave e a sua rede aleatória equivalente (rede com o mesmo número de vértice e grau médio), conforme Tabela 2, verificou-se o fenômeno mundo pequeno (*Small-World*):

Tabela 2: Índices verificados para a maior componente da rede semântica baseada em palavras-chave e a sua rede aleatória equivalente. Fonte: Nascimento et al. (2016).

Rede	Propriedade	Valor	Propriedade	Valor
Palavras-chave	$n = V $	669	Maior Componente (%)	98,95
	$m = E $	5910	$\langle k \rangle$	8,8
	Δ	0,013	C_{ws}	0,74
	L	3,06	D	6
	$n = V $	669	Maior Componente (%)	100

Rede aleatória equiv.	$m = E $	3002	$\langle k \rangle$	8,8
	Δ	0,013	C_{ws}	0,012
	L	3,02	D	6

A topologia da rede semântica baseada em palavras-chave (Figura 1), verificada por meio da distribuição de graus realizada, indica um padrão que segue uma lei de potência, da forma $P(k) \sim k^{-\gamma}$.

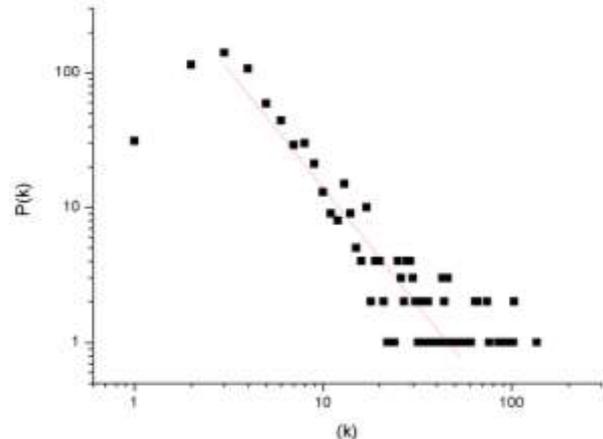


Figura 2 - Distribuição de graus da rede semântica (com $\gamma = 1,73784$ e o ajuste $R^2 = 0,85309$). Fonte: Nascimento et al. (2016).

Assim a rede semântica da Figura 1 apresentou as topologias de livre de escala, conforme a lei de potência sugerida e também, uma rede mundo pequeno. Ambas as topologias são possíveis, haja vista não serem excludentes [1,4]. Ao compararmos a pré-formatação das palavras-chave conforme Nascimento et al. [1] e a proposta deste artigo, verificamos que as regras apresentadas por Pereira et al. [4] para títulos de trabalhos científicos (dependendo da interpretação) podem resultar, quando aplicadas a palavras-chave, em novos vértices na rede. A título de exemplo, consideremos o seguinte conjunto de palavras-chave pertencentes a três dissertações do Ensino de Física no Brasil, no período de 1996-2006:

D1: *Ciência e Arte, Física e Cultura, Pintura; Física Moderna;*

D2: *Ciência e Religião, Ciência e Espiritualidade, Filosofia, Psicologia, Física Moderna;*

D3: *Atualização Curricular, Física Moderna, Mecânica Quântica.*

De forma didática, apresentamos na Figura 3 os vértices pré-formatados para a construção das redes semânticas:

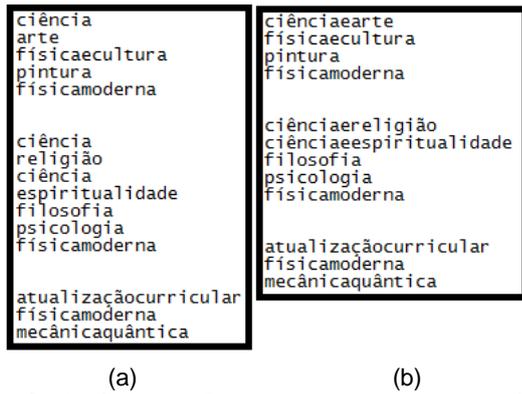


Figura 3 – (a) Pré-formatação realizada em Nascimento et al. [1]. (b) pré-formatação realizada para a construção da rede semântica proposta nesta pesquisa.

A rede semântica baseada nas palavras-chave, de acordo com as informações na Figura 3a é apresentada na Figura 4:

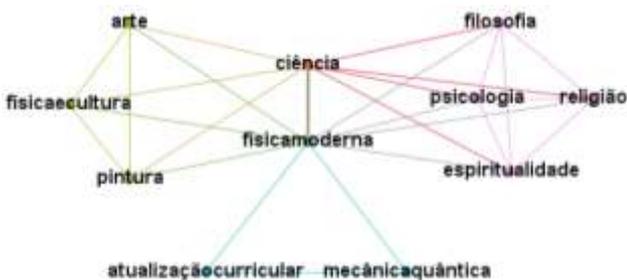


Figura 04: Rede semântica baseada em palavras-chave com o método adaptado e apresentado em Nascimento et al. [1]. Fonte: Dos autores.

Conforme as regras de Pereira et al. [4] e adaptadas em Nascimento et al. [1], destacando as regras 3, 6, 9 e 10 dispostas no Quadro 1, propomos a seguinte regra às redes semânticas baseadas em palavras-chave:

R12: As palavras que constituem as palavras-chave devem se tornar uma só.

Assim, com este destaque nas regras de pré-processamento, teremos os vértices presentes na Figura 3b, originando a rede semântica da Figura 5:

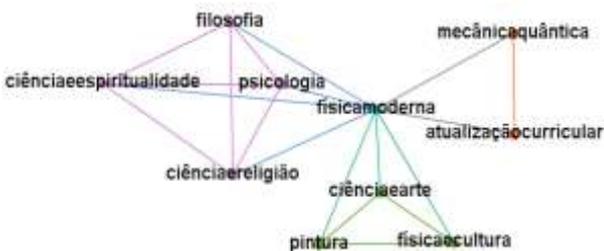


Figura 5: Rede semântica baseada em palavras-chave com as observações complementares a Nascimento et al. [1].

Comparando as Figuras 3, 4 e 5 observamos que, com a proposição da regra citada anteriormente

(R12), as cliques originadas de palavras-chave são constituídas fidedignamente, em que cada vértice passa a se tornar exatamente cada palavra-chave do conjunto de palavras utilizadas numa obra científica. Desta forma a rede semântica baseada em palavras-chave do Ensino de Física Brasileiro no período de 1996-2006 passa a ser a representada na Figura 6:

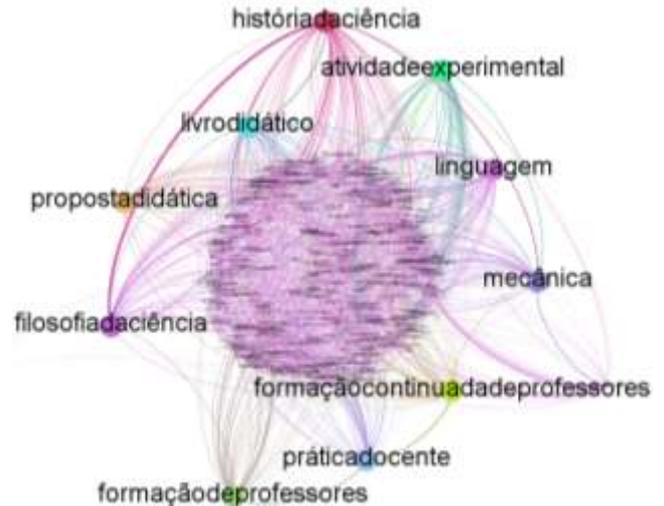


Figura 6: Rede semântica baseada em palavras-chave de dissertações e teses em Ensino de Física no Brasil. Fonte: Dos autores.

Na Figura 6 estão destacados os dez vértices com maiores centralidades de grau. As propriedades verificadas para a rede semântica da Figura 6 estão presentes na Tabela 3:

Tabela 3: Propriedades pertencem à Rede Semântica de palavras-chave e sua maior componente. Fonte: Dos autores.

Rede	Propriedade	Valor	Propriedade	Valor
Palavras-chave	$n = V $	634	Maior Componente (%)	98,9
	$m = E $	2708	$\langle k \rangle$	8,543
	Δ	0,013	C_{ws}	0,073
	L	3,061	D	6
Maior Comp.	$n = V $	627	Maior Componente (%)	100
	$m = E $	2703	$\langle k \rangle$	8,622
	Δ	0,014	C_{ws}	0,733
	L	3,061	D	6

Comparando os dados da maior componente da rede semântica baseada em palavras-chave dos trabalhos *Stricto Sensu* do Ensino de Física Brasileiro, com a sua rede aleatória equivalente (rede aleatório com o mesmo número de vértices e grau médio), conforme Watts e Strogatz [8], percebemos que a rede (Figura 6), apresenta também o fenômeno mundo pequeno, conforme Tabela 4:

Tabela 4: Índices verificados para a maior componente da rede semântica baseada em palavras-chave e a sua rede aleatória equivalente. Fonte: Dos autores.

Rede	Propriedade	Valor	Propriedade	Valor
Maior Componente	$n = V $	627	Maior Componente (%)	100
	$m = E $	2703	$\langle k \rangle$	8,622
	Δ	0,014	C_{ws}	0,733
	L	3,061	D	6
Rede Aleatória Equivalente	$n = V $	627	Maior Componente (%)	100
	$m = E $	2703	$\langle k \rangle$	8,543
	Δ	0,014	C_{ws}	0,013
	L	3,193	D	6

Em relação a topologia, a distribuição de graus da rede semântica da Figura 6, sugere uma lei de potência, da forma $P(k) \sim k^{-\gamma}$:

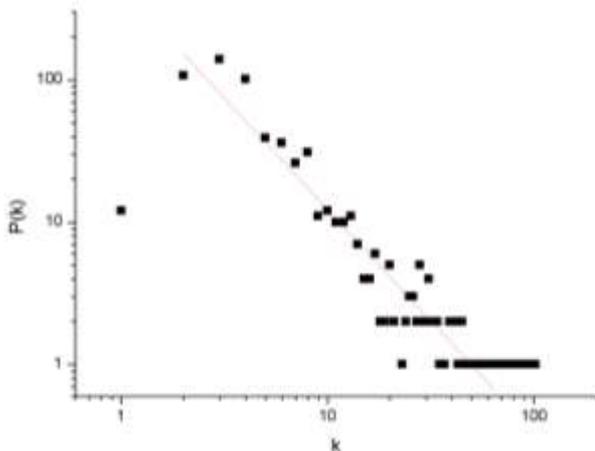


Figura 7 - Distribuição de graus da rede semântica (com $\gamma = 1,5667$ e o ajuste $R^2 = 0,9087$). Fonte: Dos autores.

Em relação aos vértices com maiores centralidades de grau nas redes das Figuras 1 e 6, para compararmos os resultados, apresentamos no Quadro 2:

Quadro 3: Vértices com maiores centralidades de grau. Fonte: Dos autores.

Fonte de dados	Nascimento et al. (2016)		Rede semântica proposta – Figura 6	
	Vértice (V)	Grau (k)	Vértice (V)	Grau (k)
1	História da Ciência	102	História da Ciência	102
2	Livro didático	97	Livro didático	93
3	Atividade experimental	89	Atividade experimental	87
4	Linguagem	85	Linguagem	81
5	Formação continuada de professores	76	Formação continuada de professores	75
6	Formação de professores	74	Formação de professores	70

7	Proposta didática	74	Prática docente	69
8	Filosofia da ciência	67	Proposta didática	64
9	Concepção do professor	66	Filosofia da ciência	61
10	Mecânica	64	Mecânica	60

O Quadro 4 apresenta alguns assuntos de física verificados nas redes semânticas das Figuras 1 e 6. Mostra os dez primeiros vértices com maiores centralidades de grau:

Quadro 4: Vértices com maiores centralidades de grau. Fonte: Dos autores.

Fonte de dados	Nascimento et al. (2016)		Rede semântica proposta – Figura 6	
Ordem	Vértices (V)	Grau (k)	Vértices (V)	Grau (k)
1	Mecânica	64	Mecânica	60
2	Física moderna e contemporânea	53	Astronomia	45
3	Astronomia	47	Eletromagnetismo	42
4	Eletromagnetismo	44	Energia	33
5	Óptica	33	Eletricidade	32
6	Gravitação	26	Óptica	31
7	Termodinâmica	25	Física Moderna	31
8	Mecânica Quântica	14	Física Térmica	29
9	Relatividade Restrita	12	Gravitação	26
10	Cosmologia	9	Termodinâmica	25

O Quadro 4 apresenta alguns assuntos de física verificados nas redes semânticas das Figuras 1 e 6. Ele nos mostra os dez primeiros vértices com maiores centralidades de grau. Desta forma, a proposta metodológica apresentada na presente pesquisa, indica que os vértices mais importantes por meio de suas conexões com outros vértices sofrem uma forte alteração na rede semântica. Conforme comparação dos Quadros 3 e 4. Cosmologia em ambas as redes, é destacado como um dos assuntos menos explorados no Ensino de Física nacional, no período investigado. Na rede semântica da Figura 6 apresentou ($k=8$).

4. Considerações finais

Utilizando a Teoria dos Grafos e de Redes, por meio da construção e análise de duas redes sociais

e complexas, realizamos um estudo na metodologia de construção das redes semânticas baseadas em palavras-chave pertencentes às dissertações e teses defendidas no Brasil no período de 1996-2006.

Por meios dos índices de estatística básica da teoria de Redes Complexas e da métrica de Centralidade de Grau pertencente às Redes Sociais, verificamos o impacto que a metodologia apresentada impõe na rede semântica resultante. Ao analisarmos a maior componente de ambas as redes (representação do sistema) o número de vértices (n) na rede proposta e construída na presente pesquisa apresentou o valor de 627 palavras-chave o que satisfaz na íntegra a quantidade total do sistema. O número de arestas (E) correspondeu a aproximadamente a metade do valor inicial pesquisado, resultando em 2703 conexões, satisfazendo as ligações entre os vértices no sistema. Em relação ao grau médio ($\langle k \rangle$), densidade (Δ), coeficiente de aglomeração (C_v) e coeficiente de aglomeração médio (C_{ws}) caminho mínimo médio (L) e diâmetro (D), os valores corresponderam ser da mesma ordem.

Em relação a topologia, a distribuição de graus das redes semânticas nos sugere uma lei de potência, da forma $P(k) \sim k^{-\gamma}$ e, havendo indícios de serem redes *Scale Free* (livres de escala). Mesmo o valor do coeficiente γ encontrado para ambas as redes ser menor do que o proposto por Barabási e Albert [9] ($2,1 \leq \gamma \leq 4$), verificamos em literatura científica redes semânticas com o valor de γ , como em Fadigas et al. [2] $\gamma \leq 2,2$ e $\gamma < 2$; na pesquisa de Seyed-allaei et al. [10]. Ao utilizarmos o método proposto de Watts e Strogatz [8], verificamos que as redes também apresentam características de serem *Small-World* (mundo pequeno). É válido destacar que uma rede complexa, pode ser ao mesmo tempo *Scale Free* e *Small World*. Uma consequência da alteração do número de vértices (n) e a radical alteração no valor das arestas (E), apresentaram um impacto direto na centralidade de grau dos vértices. Assim, ocorreram alterações nos vértices mais importantes – devido a centralidade de grau – e, mais notoriamente, nos vértices que correspondem a assuntos abordados no Ensino de Física nacional. A rede semântica proposta nesta pesquisa, assim como em Nascimento et al. [1] nos faz indicar que apresenta características como o crescimento e a adesão preferencial. Há maior probabilidade de ocorrerem novas palavras-chave com as palavras do Quadro 4 do que com as do Quadro 3. Pelas palavras-chave estudadas ainda verificamos que não se objetivou dissertações e teses voltadas para metodologias de assuntos no Ensino de Física e sim a formação de físicos pesquisadores em temáticas voltadas à educação em ciências/Ensino de Física. Por

fim, a técnica apresentada à construção de redes semânticas de palavras-chave está em consonância com Pereira et al. [4] e, pode ser utilizada para qualquer estudo que faça utilização da Teoria de Redes, tendo como vértices as referidas palavras-chave.

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-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.
- [2] 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.
- [3] 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.
- [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] SALEM, S.; KAWAMURA, M. R. D. Ensino de Física no Brasil: catálogo analítico de dissertações e teses (1996-2006). São Paulo: Instituto de Física da USP / PROFIS, 2009. 243 p.
- [6] GROSS, J. L., YELLEN, J. Graph theory and its applications. CRC press, Boca Raton FL USA 2005.
- [7] CALDEIRA, S. M. G. "Caracterização da Rede de Signos Linguísticos: 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, 2005.
- [8] WATTS, Duncan J.; STROGATZ, Steven H. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 6684, p. 440-442, 1998.
- [9] BARABÁSI, A. L. E ALBERT, R.. "Emergence of scaling in random networks". *Science*, n. 286, pp. 509-512, 1999.
- [10] SEYED-ALLAEI, H, BIANCONI & M, G. MARSILI. Scale-free networks with an exponent less than two. *Phys. Rev. E* 73, 046113, 2006.