

Redes semânticas complexas baseadas em palavras-chave: uma análise do Ensino de Física Brasileiro por meio de *Time-Varying Graphs*

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

¹ Departamento de Modelagem Computacional, CIMATEC, Salvador/BA, Brasil.

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

* e-mail: jeffersonascimento@gmail.com

Resumo

As análises de redes complexas podem ocorrer por meio de um percurso metodológico alicerçado em índices estatísticos estáticos e/ou dinâmicos. Este artigo tem como objetivo apresentar o método de construção e as análises de redes semânticas baseadas em palavras-chave. O espaço amostral escolhido como material para análise correspondeu às dissertações e teses defendidas em Ensino de Física no Brasil, entre os anos de 1972 a 2006. Construímos e analisamos as redes de palavras-chave, por meio de um modelo computacional evolutivo no tempo, utilizando um *Time-Varying Graphs* (TVG). Os resultados indicaram as variações temporais pelas janelas de tempo, anualmente, fornecendo indícios de como ocorreu a evolução da área de Ensino de Física no Brasil. Por fim, as análises dinâmicas pertencentes à Teoria e Ciências das Redes, apresentaram-se como uma adequada ferramenta para indicar a evolução do Ensino de Física nacional e qualquer área do conhecimento.

Palavras-chave: Teoria e Ciência de Redes, Sistemas Complexos, Séries Temporais, Ensino de Física, Redes Semânticas.

Abstract

Neste espaço colocar o texto correspondente em inglês. Este item é obrigatório. The analysis of complex networks can occur through a methodological path based on static and / or dynamic statistical indexes. This article aims to present the construction method and the analysis of semantic networks based on keywords. The sample space chosen as material for analysis corresponded to the dissertations and theses defended in Physics Teaching in Brazil, between the years 1972 to 2006. We built and analyzed the keyword networks, using a computational model that evolved over time, using a *Time-Varying Graphs* (TVG). The results indicated the temporal variations by the time windows, annually, providing indications of how the evolution of the area of Physics Teaching in Brazil occurred. Finally, the dynamic analyzes belonging to Theory and Science of Networks, presented themselves as an adequate tool to indicate the evolution of national Physics Education and any area of knowledge.

Keywords: Network Theory and Science, Complex Systems, Time Series, Physics Teaching, Semantic Networks.

1. Introdução

A investigação dos sistemas complexos ocorre por meio do formalismo matemático de sistemas dinâmicos, através de equações diferenciais, equações por meio de diferenças, mapas logísticos, autômatos celulares, redes dinâmicas (teoria de redes), etc. (ibidem), cuja ferramenta fundamental é a utilização de

recursos computacionais, conforme Moyano [1]; Teixeira [2]; Cunha [3], Nascimento et al. [4]). Então, por conta das dificuldades encontradas e relacionadas aos estudos desses sistemas que se apresentam “[...] fora do equilíbrio, a única ferramenta disponível, na prática, é o computador” [5].

O exposto até aqui não corresponde uma definição de sistemas complexos, haja vista não

encontrarmos na literatura científica uma única concepção teórica sobre o que ele é. Contudo, os supracitados autores indicam as seguintes propriedades para um sistema complexo: partes que se relacionam entre si; interação com o meio; adaptação ao meio; tratamento da informação em vários níveis; ordem emergente; propriedades coletivas emergentes, resultando em novos comportamentos causados pela interação entre as partes; quebra de ergodicidade; criticalidade autoorganizada, caos, estrutura fractal, seguem uma lei de potência, dentre outras.

Como mencionado anteriormente, uma forma de investigação dos sistemas complexos ocorre por meio da teoria de redes. Assim, para Pereira [6] as redes sociais e complexas apresentam as seguintes características:

[...] possuem propriedades não triviais, inexistentes em modelos mais simples. Essas redes são sistemas complexos que, em geral, envolvem inúmeros elementos organizados em estruturas que podem existir, ou coexistir, em diferentes escalas. Seus processos de ação e de organização não são usualmente descritos por regras simples ou redutíveis a apenas um nível explanatório. Frequentemente, suas características principais emergem de interações entre suas partes constituintes e não podem ser previstas a partir de uma compreensão isolada de cada uma destas partes. Nesse sentido, deparamo-nos com a especialização de técnicas e métodos de cálculo para propriedades emergentes (e.g. densidade, coeficiente de aglomeração, caminho mínimo, distribuição de graus, centralidade e prestígio, etc.) de redes sociais e complexas. O estudo desses índices favorece a análise, interpretação e resolução de problemas em diversas áreas.

Nesse contexto, esta pesquisa tem como objetivo investigar as palavras-chave como elementos fundamentais para identificar um padrão no desenvolvimento do Ensino de Física no Brasil. Para tanto, observaremos desde a primeira dissertação considerada na área até a última tese (conforme base de dados oficial) entre 1972-2006. Assim, analisamos as interações das palavras-chave das dissertações e teses do Ensino de Física Brasileiro, em busca do sistema complexo que represente a sua evolução temporal.

A realização da presente pesquisa está alicerçada na teoria dos grafos estáticos, grafos dinâmicos e pela teoria e ciência de redes. Construiremos as redes semânticas estáticas baseadas em palavras-chave das dissertações e teses em Ensino de Física, calcularemos os índices das redes complexas para verificarmos o supracitado padrão evolutivo do Ensino de Física Brasileiro. Para atingirmos este objetivo realizaremos a construção das redes semânticas complexas dinâmicas, por meio de um *Time-Varying Graphs* (TVG).

2. Metodologia

A base de dados de dissertações e teses utilizadas nesta pesquisa, corresponde à citada em Nascimento, Pereira e Moret [7]. Para que pudéssemos realizar as análises de redes semânticas por meio de um estudo da evolução temporal das redes baseadas em palavras-chave, utilizamos o formalismo referente aos Grafos que variam no Tempo, *Time-Varying Graphs* (TVG). Para considerarmos o caráter dinâmico nas redes semânticas (há alterações nos vértices e em suas interações), torna-se necessário a representação por meio de grafos cujos vértices e arestas sejam não-fixos, conforme Santana [8] e Cunha [3]). Um TVG

corresponde a um grafo estático $G = (V, \mathcal{E})$ onde são acrescentadas as informações ou parâmetros que correspondem às funções atemporais (conjuntos atemporais) [3]. Essas funções ou conjuntos atemporais são a função de presença (Y), a função de latência (ζ) e o tempo de vida do sistema (Γ).

Assim, em conformidade com Casteigts et al. [9], Santana (2012) e Cunha [3] um TVG é representado pela quintupla $G = (V, \mathcal{E}, Y, \zeta, \Gamma)$, em que: V e \mathcal{E} correspondem, respectivamente, o número de vértices e arestas do sistema. A função de presença

$Y : \mathcal{E} \times \Gamma \rightarrow \{0, 1\}$ garante a existência de uma dada aresta em um dado instante de tempo t . A função de latência (ζ) indica quanto tempo se necessita para que uma aresta esteja disponível em um instante $t \in \Gamma$, ou seja, é o tempo necessário para estabelecer a conexão entre dois vértices, em um dado instante t . A função $\Gamma \subset N$ representa o tempo de vida do sistema.

Alicerçados nos supracitados autores e diante das informações anteriores, temos para a presente pesquisa do Ensino de Física Brasileiro:

- $V = \{v_1, v_2, v_3, \dots, v_n\}$ correspondem aos conjuntos de vértices das redes semânticas, ou seja, as palavras pertencentes as cliques de palavras-chave;
- $\mathcal{E} = \{e_1, e_2, e_3, \dots, e_n\}$ representam os conjuntos de arestas, isto é, as conexões realizadas entre palavras-chave;
- Em relação a função presença (Y), inspirados em Casteigts et al. [9], Silva et al. [10] e Cunha [3], o TVG baseado em palavras-chave corresponde, respectivamente, a um conjunto de grafos

estáticos da forma $G^{[t_i, t_j]} = (V, \mathcal{E}^{[t_i, t_j]})$. Assim,

$$\forall e \in \mathcal{E}, \quad e \in \mathcal{E}^{[t_i, t_j]} \Leftrightarrow \exists t \in [t_i, t_j], Y(e, t) = 1$$

$e \in \mathcal{E}^{[t_i, t_j]}$. Teremos então a seguinte representação:

$$G = \{G^{[t_i, t_j]}, G^{[t_{i+1}, t_{j+1}]}, G^{[t_{i+2}, t_{j+2}]}, G^{[t_{i+3}, t_{j+3}]}, \dots, G^{[t_{i+n}, t_{j+n}]} \}$$

- A função de latência (\mathcal{S}) é o tempo necessário para que duas palavras nas redes baseadas Palavras-chave possam se conectar. Para esta pesquisa, a função latência é constante, não sendo computada nas análises.
- O tempo de vida do sistema corresponde a $\Gamma = \{t_1, t_2, t_3, \dots, t_{i+1}, \dots, t_n\}$, com $n=34$. O que significa dizer que o TVG baseado em Palavras-chave, t_i corresponde a um intervalo temporal de 1 ano, ou seja, $|\Gamma| = 34$ anos.
- As janelas temporais (τ) são definidas da seguinte forma $\tau_i = [t_i, t_{i+1}]$, então, o TVG em função destas janelas, pode ser representado da seguinte maneira: $G = \{G^{\tau_1}, G^{\tau_2}, G^{\tau_3}, \dots, G^{\tau_{34}}\}$.

Desta forma o TVG do Ensino de Física Nacional corresponde ao conjunto dos subgrafos estáticos das redes baseadas em palavras-chave de um ano de dissertações e teses defendidas. Desta forma, para cada ano t , iniciado em 1972 e finalizado em 2006, há uma rede semântica baseada em palavras-chave dos trabalhos *Stricto Sensu* em Ensino de Física. Conforme Casteigts et al. [9], Silva et al. [10], Cunha [3] e Barjon et al. [11], temos então, $G(V_t, \mathcal{E}_t, t) = G(V_1 + V_2 + V_3 + \dots + V_{34}, \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 \dots + \mathcal{E}_{34})$

3. Discussão dos resultados

Por meio da metodologia apresentada na sessão anterior, construímos as redes semânticas complexas baseadas em palavras-chave de dissertações e teses do Ensino de Física Brasileiro. Desta forma, realizamos as análises dinâmicas, por meio de grafos variantes no tempo. Para o período correspondente à base de dados, a rede semântica complexa estática do Ensino de Física no Brasil é a correspondente na Figura 1:

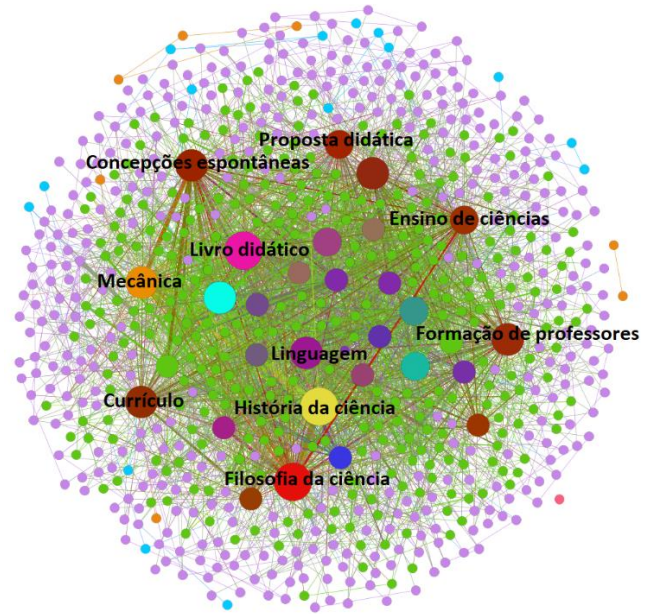


Figura 1: Rede semântica complexa estática completa baseada em palavras-chave (1972-2006). Fonte: Nascimento, Pereira, Moret [7].

Ressaltamos que os índices estatísticos utilizados em relação à rede complexa da Figura 1, são os mesmos presentes em Nascimento, Pereira, Moret [7]. Correlacionamos os resultados das redes semânticas estáticas, como a da Figura 1, com as redes dinâmicas por meio do *Time-Varying Graphs*.

Os índices utilizados correspondem aqueles denominados de atemporais ou índices clássicos de redes, conforme Santoro et al. [12], Cunha [3], Cunha et al. [13] e Nascimento [14]: número de vértices (n), grau médio ($\langle k \rangle$) número de arestas (m), coeficiente de aglomeração (C_{ws}), caminho mínimo médio (L) e densidade (Δ):

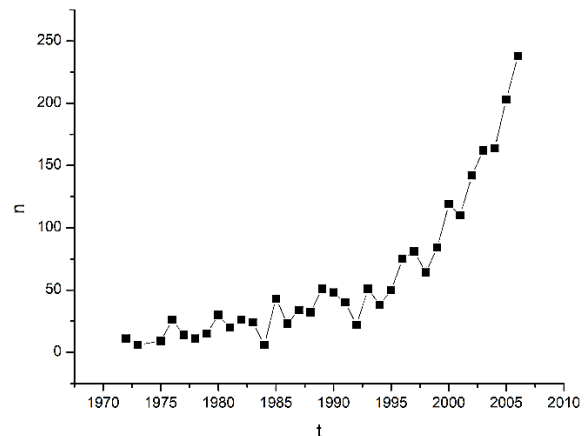


Figura 2 - TVG do coeficiente n para o Ensino de Física Brasileiro (1972-2006). Fonte: Adaptado de Nascimento [14].

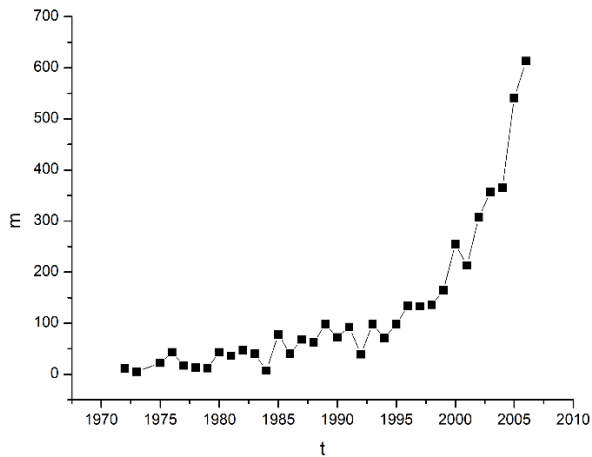


Figura 3 - TVG do coeficiente m para o Ensino de Física Brasileiro (1972-2006). Fonte: Adaptado de Nascimento [14]

Analogamente aos resultados encontrados com Nascimento, Pereira e Moret (2020), conforme Figura 2 e Figura 3 as séries temporais indicaram que tanto o número de vértices quanto o de arestas são proporcionais entre si, para o TVG baseado em palavras-chave. Fato este relacionado ao número crescente de pesquisas na área de Ensino de Física, ocorrendo com o mesmo padrão. A seguir, apresentamos o TVG dos índices número de vértices grau médio ($\langle k \rangle$), densidade (Δ), coeficiente de aglomeração (C_{ws}), caminho mínimo médio (L) e diâmetro (D):

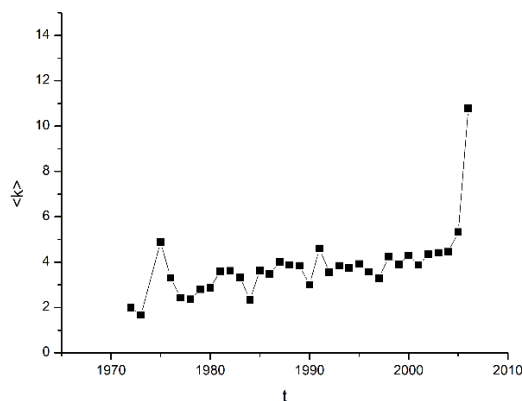


Figura 4 - TVG do coeficiente $\langle k \rangle$ para o Ensino de Física Brasileiro (1972-2006). Fonte: dos autores

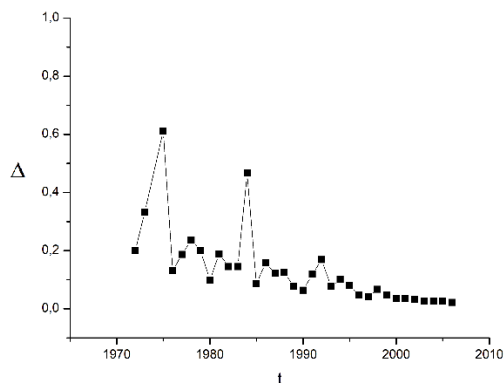


Figura 5 - TVG do coeficiente Δ para o Ensino de Física Brasileiro (1972-2006). Fonte: Adaptado de Nascimento [14]

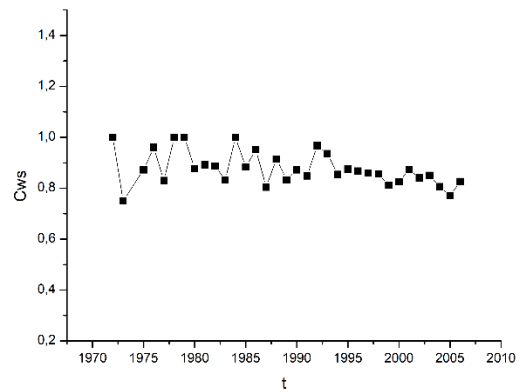


Figura 6 - TVG do coeficiente C_{ws} para o Ensino de Física Brasileiro (1972-2006). Fonte: Adaptado de Nascimento [14]

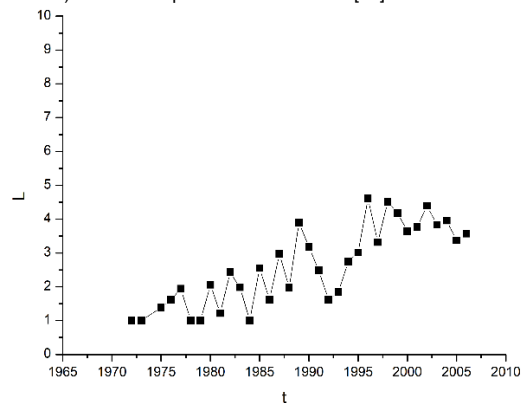


Figura 7 - TVG do coeficiente L para o Ensino de Física Brasileiro (1972-2006). Fonte: Adaptado de Nascimento [14]

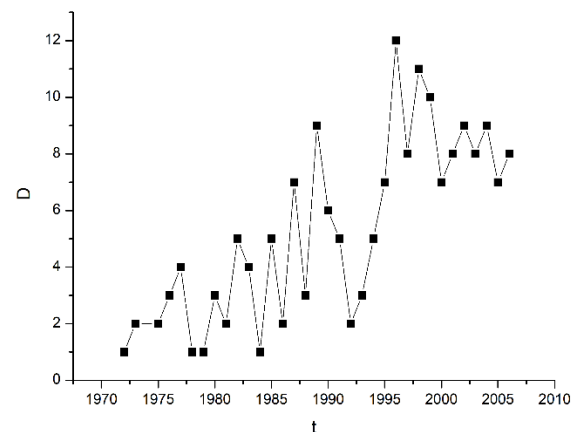


Figura 8 - TVG do coeficiente D para o Ensino de Física Brasileiro (1972-2006). Fonte: dos autores

Como citado anteriormente, as Figuras 2-8 apresentam total coerência com os resultados obtidos pelas análises de TVG baseados em Títulos, conforme Nascimento, Pereira e Moret [15]. Todas as métricas do *Time-Varying Graphs* aqui apresentadas, número de vértices (n), número de arestas (m), grau médio ($\langle k \rangle$), densidade (Δ), coeficiente de aglomeração (C_{ws}), caminho mínimo médio (L) e diâmetro (D) apresentaram um comportamento semelhante na década final do espaço amostral. Entre o período de 1972-1995, percebemos uma flutuação nas séries temporais, estabilizando na década final, 1996-2006.

Um indício deste comportamento representado pelo TVG é a alta concentração de pesquisas em emáticas voltadas ao Ensino de Física que ocorreram neste período, tornando as redes, de forma singular, com esta característica evolução. Para exemplificar este raciocínio, destacamos na presente análise, escolhidos de forma aleatória, as métricas do grau médio ($\langle k \rangle$) e o coeficiente de aglomeração (C_{ws}).

Em 2006, último ano da série temporal $\langle k \rangle = 10,79$. Este valor indica, em média, quão elevadas foram as conexões realizadas pelas palavras-chave contidas em dissertações e teses distintas. O $C_{ws} = 82,6\%$, ratifica o resultado anterior e demonstra o quão conectada e eficiente é, em termos de difusão da informação, a rede semântica complexa que se apresentou em 2006. Apesar de termos tomado 2006 como exemplo, reiteramos que este é o padrão que veio evoluindo no Ensino de Física Nacional, iniciando em 1972 e finalizando em 2006, intervalo que corresponde ao nosso espaço amostral desta pesquisa.

Por fim, a correlação entre as redes semânticas estáticas originadas, bem como a rede complexa presente na Figura 1, que abrange todo o espaço amostral desta pesquisa, apresentou um resultado coerente com as pesquisas de Nascimento, Pereira e Moret [15]: o Ensino de Física Brasileiro evoluiu por um padrão, cuja observância correspondeu temas como História e Filosofia da Ciência; recursos metodológicos; formação de professores; psicologia cognitiva; pesquisas voltadas para o Ensino de Física tanto à Educação Básica, quando o nível superior.

4. Considerações finais

Analizamos a correlação entre as redes semânticas estáticas e que apresentam uma dinâmica de evolução temporal (*Time-Varying Graphs*) baseadas em palavras-chave das dissertações e teses do Ensino de Física no Brasil, pelo período de 1972-2006. Propomos e concretizamos o TVG do Ensino de Física nacional, o qual foi representado por uma quintupla, que além dos Vértices e as Arestas, apresentaram os conjuntos atemporais: a função de presença (Υ), a função de latência (\mathcal{L}) e o tempo de vida do sistema (Γ).

As métricas atemporais ou índices clássicos de redes, que suscitaram séries temporais e o *Time-Varying Graphs* aqui apresentadas, como, o número de vértices (n), número de arestas (m), grau médio ($\langle k \rangle$), densidade (Δ), coeficiente de aglomeração (C_{ws}), caminho mínimo médio (L) e diâmetro (D) apresentaram uma tendência de estabilização na década final, 1996-2006. Estes fatos estão relacionados a alta conectividade presentes na rede pelas diversas temáticas, pelas curtas distâncias entre elas e a presença de *hubs*.

Por fim, em todo o período analisado foi possível concluir os tópicos/categorias estabelecidas pela observância dos *hubs* que levaram a observarmos os indícios qualitativos da evolução do Ensino de Física Nacional, em que destacamos/elencamos: História e Filosofia da Ciência; recursos metodológicos; formação de professores; psicologia cognitiva; pesquisas voltadas para o Ensino de Física tanto à Educação Básica, quando o nível superior.

5. Agradecimentos

Jefferson Nascimento agradece a oportunidade de ter realizado pós-doutoramento no departamento de modelagem computacional no CIMATEC/BA. Marcelo A. Moret agradece ao CNPq pelo suporte financeiro parcial oriundo de sua bolsa de Produtividade em Pesquisa (No. 305291/2018-1).

6. Referências

- [1] MOYANO, L. G. Mecânica estatística não-extensiva em sistemas complexos: fundamentos dinâmicos e aplicações. Tese de Doutorado, Centro Brasileiro de Pesquisas Físicas. Rio de Janeiro, 2006.
- [2] TEIXEIRA, G. M. Redes semânticas baseadas em discursos orais: Uma proposta metodológica baseada na psicologia cognitiva utilizando redes complexas. Dissertação (Mestrado Interdisciplinar em Modelagem Computacional). Fundação Visconde de Cairu, Salvador, 2007.
- [3] CUNHA, M.V. Redes semânticas baseadas em títulos de artigos científicos. 127f. Dissertação (Mestrado em modelagem computacional e tecnologia industrial) – Faculdade de Tecnologia Senai CIMATEC, Salvador, 27 nov. 2013.
- [4] NASCIMENTO, Jefferson Oliveira do; Pereira-Guizzo, Camila de Sousa; Monteiro, Roberto Luiz Souza ; Moreira, Davidson Martins; Moret, Marcelo Albano; Pereira, Hernane Borges De Barros . The postgraduate Brazilian studies in Physics Teaching using Complex Network. In: 6th International Conference on Nonlinear Science and Complexity, 2016, São José dos Campos - SP. v. 1. doi:10.20906/CPS/NSC2016-0010
- [5] OLIVEIRA, P. M. C. Autômatos Celulares. In: NUSSENZVEIG, H. M. [Org.]. Complexidade & caos. Rio de Janeiro: UFRJ/ COPEA, 2003.
- [6] PEREIRA, H. B. B. Redes Sociais e Complexas: Aplicações em Difusão do Conhecimento. Academia de Ciências da Bahia: Memória, v. 3, p. 39-47, 2013.
- [7] NASCIMENTO, J. O. do; HERNANE, B. B. P; 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.

[8] SANTANA, A. Caracterização da jornada máxima em redes dinâmicas. 307 p. Dissertação (Mestrado) | Universidade Federal da Bahia, Salvador, 2012.

[9] CASTEIGTS, A.; FLOCCHINI, P.; QUATTROCIOCCHI, W.; SANTORO, N. Time-varying graphs and dynamic networks. In: ADHOC-NOW. [S.l.: s.n.], p. 346-359, 2011.

[10] SILVA, B. B. M. et al. Statistical characterization of an ensemble of functional neural networks. Social Networks European Physical Journal B, v. 392, p. 85{358, 2012.

[11] BARJON, M et al. Testing temporal connectivity in sparse dynamic graphs. arXiv preprint arXiv:1404.7634, 2014.

[12] SANTORO, N. On the temporal analysis of scientific network evolution. In: CASoN. p. 169-174, 2011.

[13] CUNHA, M. V. et al. (Ed.). Redes de títulos de artigos científicos variáveis no tempo. [S.l.]: In: Anais II Workshop on Social Network Analysis and Mining (BraSNAM 2013) e XXXIII Congresso da Sociedade Brasileira de Computação (CSBC 2013), 2013. 12 p.

[14] NASCIMENTO, J. O. do. Redes sociais e complexas: redes semânticas do Ensino de Física Brasileiro. 123f. Tese (Doutorado em modelagem computacional e tecnologia industrial) – CIMATEC, Salvador, 28 jun. 2019.

[15] NASCIMENTO, J. O. do; PEREIRA, H. B. B.; MORET, M.A. Time-Varying Graphs e séries temporais: análise do Ensino de Física nacional por meio de redes semânticas complexas In: XXIII Encontro Nacional de Modelagem Computacional e o XI Encontro de Ciência e Tecnologia de Materiais, 2020, Palmas/TO. Anais do XXIII Encontro Nacional de Modelagem Computacional e o XI Encontro de Ciência e Tecnologia de Materiais [recurso eletrônico]. 2020. v.1. p. 2014 - 2023