

## **A entropia da informação em redes semânticas de títulos variáveis no tempo.**

**Marcelo do Vale Cunha (Doutorando - MCTI), [celaocunha@gmail.com](mailto:celaocunha@gmail.com);**

**Hernane Borges de Barros Pereira (Orientador - MCTI), [hbbpereira@gmail.com](mailto:hbbpereira@gmail.com);**

**Faculdade SENAI CIMATEC**

Palavras Chave: *Redes Variáveis no Tempo, Rede de títulos, Entropia da Informação*

### **Introdução**

As publicações de pesquisas em artigos científicos constituem um meio de comunicação científica. Neste contexto, o sistema formal de comunicação é o periódico científico, que vem sendo largamente estudado nas últimas décadas, por terminologias, funcionalidades e categorização (Garvey, 1979; Ziman, 1979; Miranda, 1996; Vanz, 2010).

As redes semânticas de títulos (*RST*) podem modelar esta comunicação científica através do vocabulário contido em títulos de um certo conjunto de publicações, como exemplo: um número de um periódico científico ou um período de publicações.

Pereira et al (2011) estabeleceram as regras de criação para estas redes e investigaram as propriedades topológicas para 15 periódicos científicos de alto impacto. Cunha et al (2013) propuseram uma modelagem para uma *RST* como um grafo variável no tempo (*TVG*). Eles investigaram o efeito de memória nas séries temporais dos principais índices de redes. Pereira et al (2016) estudaram a evolução da densidade durante a construção de *RST* como indicador de diversidade de conceitos de periódicos científicos.

Este trabalho investiga a diversidade do vocabulário da revista *Nature* ao longo de 10 anos, utilizando a mesma modelagem dos trabalhos supracitados para construção da *RST* como um *TVG*, acrescida do cálculo da entropia da informação das redes pertencentes às janelas temporais do *TVG*.

As redes são construídas através de uma janela bimensal de publicações que avança semana a

semana, desde 1º de Janeiro de 1999 a 19 de dezembro de 2008.

Para cada janela temporal é calculado a frequência de aparição dos vértices e arestas da rede. Estas frequências são inseridas na equação da Entropia de Shannon (Shannon, 1948) e assim poder ser gerada a série temporal da Entropia.

Esta série evidencia tendências na diversidade do vocabulário da revista, ressaltando períodos onde ocorrem mudanças bruscas nos valores da entropia, podendo indicar como as temáticas variam na revista ao longo dos 10 anos.

### **Métodos e Resultados parciais**

Uma rede de títulos variáveis no tempo considera as informações temporais contidas em seus títulos, na construção da rede. Neste contexto, propomos a construção da *RST* como um Grafo Variável no Tempo (*TVG*), representado aqui pela Equação 1.

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \Gamma, \Upsilon, \varsigma\} \quad (1)$$

Na equação,  $\mathcal{G}$  é o grafo variável no tempo, que além de conter o vocabulário da revista, como o conjunto de vértices ( $\mathcal{V}$ ), os pares de palavras pertencentes a um mesmo título, como o conjunto de arestas ( $\mathcal{E}$ ), contém informações temporais, sobre o tempore vida do sistema (período coletado) ( $\Gamma$ ), a presença dos pares de palavras em cada instante ( $\Upsilon$ ) e o tempo de formação (latência) destas conexões entre os pares ( $\varsigma$ ). Esta última foi considerada constante neste trabalho. A janela temporal neste *TVG* é definida por  $\tau = \{t_i, t_{i+7}\}$  e percorre as 507 semanas do *TVG*.

**Seminário Anual de Pesquisa – 2017**

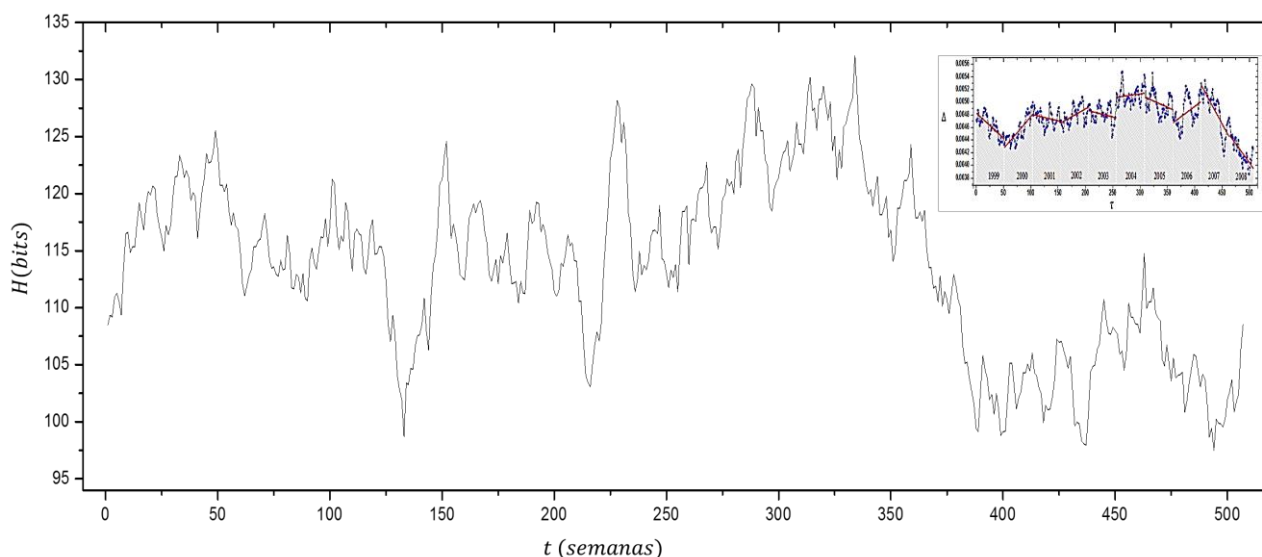
**Faculdade SENAI CIMATEC**

**Programa de Pós-Graduação em Modelagem Computacional e Tecnologia Industrial**

**Programa de Pós-Graduação em Gestão e Tecnologia Industrial**

**III Workshop de Gestão, Tecnologia Industrial e Modelagem Computacional.**

**ISSN online 2447-9640**



**Figura 1.** Série temporal da Entropia de Shannon para as arestas de cada janela do TVG. O gráfico menor mostra a variação da densidade e suas tendências para o mesmo TVG, obtidos de Cunha et al (2013).

Para cada Janela é calculado a probabilidade  $p_{jk}$  para cada aresta  $jk$ , em relação ao número total de arestas da janela  $N$ , Equação 2, onde  $j$  e  $k$  são vértices da rede.

$$p_{jk} = \frac{n_{jk}}{N} \quad (2)$$

Com isso, calcula-se a entropia de Shannon  $H_i$  para distribuição de probabilidades das arestas em cada Janela  $i$  do TVG, Equação 3.

$$H_i = \left[ \sum_{k,j=1}^n -p_{jk} \cdot \log_2 p_{jk} \right] \quad (3)$$

A Figura 1 mostra o valor da entropia de Shannon para as arestas de cada janela do TVG. Valores altos de  $H_i$  sugerem que existem uma maior diversidade de arestas diferentes na janela, ou seja, pares de palavras diferentes, do que para valores baixos de  $H_i$ .

Esta diversidade pode indicar a riqueza do vocabulário da revista, não em número de palavras, mas na capacidade delas em gerar novas associações com outras do mesmo vocabulário.

## Conclusões

A modelagem de uma rede de títulos como um grafo que varia no tempo permite uma investigação das propriedades de redes em épocas diferentes e compará-las.

O uso da Entropia de Shannon aplicado às arestas das redes viabilizou a comparação de épocas através

da diversidade de seus pares de palavras. Este método contribui para o entendimento da difusão de informação em periódicos científicos, permitindo a investigação da previsibilidade dos temas de uma revista a partir da série temporal da entropia.

## Referências

- CUNHA, M., Rosa, M. G., Fadigas, I., Miranda, J. G. V., Pereira, H. B. B. (2013). "Redes de títulos de artigos científicos variáveis no tempo". In BraSNAM - II Brazilian Workshop on Social Network Analysis and Mining, p. 1744–1755.
- GARVEY, W. D. Communication: The Essence of Science. [S.l.]: Oxford, NY: Pergamon Press, Inc., 1979. ISBN 0-08-022254-4.
- MIRANDA, D.; Pereira, M. O periódico científico como veículo de comunicação: uma revisão de literatura. *Ibict, Ci. Inf., Brasília*, v. 25, n. 3, p. 375–382, 1996.
- PEREIRA, H. B. B., Fadigas, I., Senna, V., Moret, M. (2011). "Semantic networks based on titles of scientific papers". *Physica A: Statistical Mechanics and its Applications*, 390 (6): 1192–1197.
- SHANNON, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 623–656.
- VANZ, S.; STUMPF, I. Colaboração científica: revisão teórico-conceitual. *Perspectivas em Ciência da Informação*, Scielo, v. 15, p. 42–55, 2010.