



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76

Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

## XXVIII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2024

### COMUNIDADES DE PALAVRAS EM REDES SEMÂNTICAS BASEADAS EM TEXTOS DE LIVROS DIDÁTICOS DE MATEMÁTICA

**Samara dos Santos Ferreira<sup>1</sup>; Marcos Grilo<sup>2</sup>**

1. Bolsista CNPq, Graduando em Engenharia de Computação,, Universidade Estadual de Feira de Santana, e-mail:

[samaradosantosf@gmail.com](mailto:samaradosantosf@gmail.com)

2. Orientador, Departamento de Exatas, Universidade Estadual de Feira de Santana, e-mail: [grilo@uefs.br](mailto:grilo@uefs.br)

**PALAVRAS-CHAVE:** Comunidades; redes semânticas; livros didáticos; identificação de temáticas.

### INTRODUÇÃO

A detecção de comunidades é uma abordagem eficaz para investigar a coesão de uma rede. Comunidades referem-se a grupos de vértices interconectados, caracterizados por uma concentração significativa de arestas em relação ao total da rede. Para compreender essas comunidades, é essencial considerar: a densidade da rede, que representa a proporção entre o número de arestas existentes e o número máximo possível; e a conectividade, onde uma rede é considerada conectada se, para qualquer par de vértices, existe um caminho que os conecta.

A identificação de comunidades tem diversas aplicações: em redes de citações, que pode revelar grupos de pesquisadores com interesses em áreas específicas; em redes de websites, que destaca temas abordados nas páginas; em redes metabólicas ou neurais, que representa unidades funcionais, entre outras.

Na literatura sobre análise de livros didáticos de Matemática (Sá; Júnior, 2024; Carvalho et al., 2023; Lazarini; Mendes, 2023; Silva; Seiffert Santos, 2023; Souza; Santos; Pantoja, 2023), há uma lacuna significativa quanto à investigação das relações entre as palavras que compõem seus vocabulários. Até o momento, não foram encontrados estudos que avaliem grupos densamente interconectados de palavras em livros didáticos. Este plano de trabalho de IC propôs a análise de livros didáticos de Matemática da Educação Básica por meio de redes semânticas de cliques, focando nas comunidades de palavras que constituem seus vocabulários.

O uso de redes semânticas de cliques possibilita a identificação das relações predominantes nos conteúdos dos livros didáticos, podendo oferecer aos educadores uma visão mais aprofundada sobre o que é abordado nas obras. Isso pode facilitar a seleção de materiais alinhados ao planejamento educacional. O objetivo principal deste plano de trabalho de IC é contribuir para a organização dos livros didáticos com base nas redes semânticas geradas, analisando as comunidades de palavras que emergem dos textos.

## METODOLOGIA

Neste estudo, as comunidades foram identificadas em uma rede semântica baseada em livros didáticos de Matemática, publicados antes da implementação da BNCC, período em que a separação dos conteúdos por capítulos é mais evidente. Para a conversão dos livros em LaTeX, utilizou-se o software Mathpix (Mathpix, 2024), seguido de um tratamento manual e computacional com a ferramenta TEX-2-NET (Santos e Grilo, 2023). O TEX-2-NET possibilita a construção de redes semânticas de forma integrada ao NetPal. A conversão de expressões matemáticas em palavras visaram melhorar a qualidade dos dados usados na construção da rede semântica.

A detecção de comunidades baseou-se no método Louvain (Blondel et al., 2008), implementado no software Pajek (Mrvar e Batagelj, 2024). Esse método, amplamente utilizado para análise de redes com um alto número de vértices, prioriza a modularidade, uma métrica que avalia a densidade das conexões dentro das comunidades em comparação com as conexões externas. Valores mais altos de modularidade indicam uma estrutura mais coesa de comunidades de palavras. O método foi ajustado pelo parâmetro de resolução, cujo valor influencia diretamente a granularidade das comunidades identificadas. Durante a análise, foram testados diferentes valores de resolução e a avaliação foi complementada com o índice de correlação V de Cramer, assegurando a consistência e qualidade das comunidades identificadas.

Essa abordagem permitiu uma análise refinada das comunidades presentes nos livros (Fadigas et al., 2020), destacando padrões temáticos e agrupamentos de palavras fortemente interconectadas, proporcionando uma visão detalhada sobre a estrutura de conteúdos nos materiais analisados.

## RESULTADOS E DISCUSSÃO

Inicialmente, destacamos que uma versão aprimorada desse trabalho foi submetida ao III Simpósio Brasileiro de Teoria e Ciência de Redes. Na análise do capítulo “Matrizes e Determinantes” do livro didático “Matemática: Contexto & Aplicações”, do autor Luiz Roberto Dante, volume 2, 3<sup>a</sup> edição (2016), a rede obtida é caracterizada por não ter arestas múltiplas e laços. Essa rede tem um total de 787 vértices, 12.044 arestas e com uma densidade de 0,03894.

A análise dos componentes da rede indica que 98,73% das palavras estão concentradas na maior componente, enquanto pequenas porções de palavras estão em componentes isolados. A maior componente possui um alto coeficiente de aglomeração de 0,76999 e um baixo caminho mínimo médio de 2,31817, quando comparados com uma rede aleatória equivalente, evidenciando a presença do fenômeno mundo pequeno (Watts, Strogatz, 1998). A maior componente da rede concentra quase todas as palavras, enquanto as menores componentes contém termos técnicos ou específicos, sugerindo que certas palavras aparecem de forma isolada em títulos de seções.

Quanto às palavras encontradas nos componentes menores, geralmente termos técnicos, refletem tópicos isolados que não se conectam diretamente com o restante da rede. A análise modular da maior componente identificou 4 comunidades principais, sendo que a maior, com 518 palavras, apresenta termos contextuais relacionados a situações-problemas, enquanto a segunda comunidade, com 203 palavras, é denominada

por notações matemáticas. Os resultados sugerem uma separação entre o vocabulário usado nas situações-problema contextualizadas e na sistematização da teoria matemática nos capítulos analisados.

A separação entre as comunidades reflete uma discrepância entre a abordagem contextual das situações-problema e a sistematização formal dos conceitos matemáticos nos livros analisados. Isso é particularmente notável na divisão entre a primeira e a segunda comunidades. Palavras relacionadas às situações-problema contextualizadas estão em uma comunidade, enquanto símbolos matemáticos e termos mais abstratos aparecem em outra. Os resultados encontrados sugerem que a forma como o conteúdo é introduzido, contextualizado e sistematizado no livro didático, não é completamente integrada, distanciando-se dos objetivos propostos pelos autores.

A análise da modularidade e a identificação de comunidades proporcionam uma visão mais adequada sobre aspectos da organização dos conteúdos nos livros didáticos. A presença de um vocabulário diversificado, associado a diferentes abordagens do conteúdo, reforça a necessidade de uma integração mais eficiente entre teoria e prática. O uso do método Louvain, com os devidos ajustes no parâmetro de resolução, permitiu uma visualização das comunidades, confirmando a importância de personalizar essa variável para melhorar a interpretação dos índices de redes semânticas baseadas em livros didáticos.

### **CONSIDERAÇÕES FINAIS (ou Conclusão)**

Neste trabalho, construímos redes semânticas de cliques a partir de livros didáticos utilizando a ferramenta TEX-2-NET, que, em conjunto com o NetPal, possibilita a análise de conteúdo de textos matemáticos, convertendo expressões, notações e símbolos em palavras. A rede semântica foi gerada a partir do capítulo sobre MATRIZES e DETERMINANTES do livro didático “Matemática: Contexto & Aplicações”, do autor Luiz Roberto Dante. A análise de comunidades revelou-se útil para avaliar a organização do capítulo do livro.

O ajuste do parâmetro de resolução mostrou-se fundamental para o propósito de analisar a organização do vocabulário do capítulo do livro. Valores muito baixos resultaram em uma grande comunidade que coincide com a maior componente, enquanto valores mais altos fragmentaram demais as comunidades. A escolha de um parâmetro intermediário foi feita considerando tanto os critérios do método de Louvain quanto o contexto da pesquisa.

Os resultados indicam uma separação entre as situações-problema e a sistematização matemática, o que pode sugerir uma falta de integração dos conteúdos no livro didático analisado. Como trabalhos futuros, pretendemos ampliar a base de dados, explorar outras técnicas de análise de redes e aprimorar o vocabulário de controle do TEX-2-NET, para possibilitar a investigação de outros aspectos associados aos livros didáticos.

### **REFERÊNCIAS**

SÁ, P. F. d.; JÚNIOR, A. L. L. d. Livros didáticos de matemática do 9º ano do ensino fundamental: uma análise à luz da matriz de referência da prova SAEB. Disponível em:

[https://periodicos.cfs.ifmt.edu.br/periodicos/index.php/rpd/article/view/688/754.](https://periodicos.cfs.ifmt.edu.br/periodicos/index.php/rpd/article/view/688/754)

Acesso em: 06 set. 2024..

CARVALHO, G. S. d.; MEDRADO, J. d. S.; MATTOS, R. A. L.; FONSECA, L. S. d. A abordagem da função seno no livro didático do ensino médio após a implementação da BNCC. *Revista de Ensino de Ciências e Matemática*, v. 14, n. 3, p. 1-23, 2023.

LAZARINI, L. V.; MENDES, L. O. R. Análise de possíveis problemas de progressão aritmética em livros didáticos do novo ensino médio: Um olhar por meio da resolução de problemas. *Revista Vitruvian Cogitationes*, v. 4, n. 2, p. 132-148, 2023.

SILVA, L. C. d.; SEIFFERT SANTOS, S. C. A utilização de analogias como recurso didático no ensino de matemática: uma análise estrutural em livros didáticos no ensino do conceito de função. REAMEC - Rede Amazônica de Educação em Ciências e Matemática, v. 11, n. 1, p. e23084, 2023. Disponível em: <https://periodicos.cfs.ifmt.edu.br/periodicos/index.php/rpd/article/view/688/754>.

Acesso em: 06 set. 2024.

SOUZA, A. L.; SANTOS, J. R. S.; PANTOJA, L. F. L. Sequências numéricas: um olhar sobre a abordagem feita em livros didáticos. *REVISTA FOCO*, v. 16, n. 9, p. e3083, 2023.

MATHPIX. Mathpix – OCR para matemática e LaTeX. Disponível em: <https://mathpix.com>. Acesso em: 06 set. 2024.

SANTOS, M. O.; GRILLO, M. Desenvolvimento de Ferramenta Computacional para Conversão de Representações Simbólicas em Livros Didáticos no Formato OCR PDF para TXT. *Anais dos Seminários de Iniciação Científica*, UEFS, 2023.

BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008.

MRVAR, A.; BATAGELJ, V. Pajek – program for large network analysis. Disponível em: <http://mrvar.fdv.uni-lj.si/pajek/>. Acesso em: 06 set. 2024.

FADIGAS, I.; HENRIQUE, T.; GRILLO, M.; PEREIRA, H. B. d. B. Programas de mestrado em matemática no Brasil: uma aplicação de redes para caracterizar seus títulos. *Fundacio Universitaria Balmes*, 2020.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 6684, p. 440–442, 1998