

MODELAGEM DE DADOS ABERTOS GOVERNAMENTAIS PARA IDENTIFICAÇÃO DE REDES DE ESCOLAS SIMILARES

*Government open data modeling for identifying networks of similar schools
Modelización de datos abiertos del Gobierno para identificar redes de escuelas similares*



Eneida Santana de Ávila Goulart
Doutoranda em Difusão do Conhecimento, Universidade Federal da Bahia (UFBA), Salvador, Bahia, Brasil.
Bibliotecária - Documentalista, Instituto Federal da Bahia (IFBA), Camaçari, Bahia, Brasil.
Lattes: <http://lattes.cnpq.br/9832000444934835>
ORCID: <https://orcid.org/0000-0002-0884-4756>



Roberto Luiz Souza Monteiro
Pós-doutorado em Modelagem Computacional e Tecnologia, Centro Universitário SENAI (CIMATEC), Salvador, Bahia, Brasil.
Professor, Centro Universitário SENAI (CIMATEC), Salvador, Bahia, Brasil e da Universidade do Estado da Bahia (UNEB), Alagoinhas, Bahia, Brasil.
Lattes: <http://lattes.cnpq.br/4168723811261149>
ORCID: <https://orcid.org/0000-0002-3931-5953>



Tereza Kelly Gomes Carneiro
Doutora em Difusão do Conhecimento, Universidade Federal da Bahia (UFBA), Salvador, Bahia, Brasil.
Professora, Instituto Federal da Bahia (IFBA), Camaçari, Bahia, Brasil.
Lattes: <http://lattes.cnpq.br/3305071338937583>
ORCID: <https://orcid.org/0000-0002-8081-9029>



Bruna Bomfim Lessa dos Santos
Doutora em Ciência da Informação, Universidade Federal da Bahia (UFBA), Salvador, Bahia, Brasil.
Professora, Universidade Federal da Bahia (UFBA), Salvador, Bahia, Brasil.
Lattes: <http://lattes.cnpq.br/4775068257764378>
ORCID: <https://orcid.org/0000-0003-4485-203X>

Resumo

Introdução: O objetivo deste estudo é representar as escolas similares situadas na zona urbana da Microrregião de Salvador - Bahia, de modo a identificar possíveis caminhos para colaboração entre elas, utilizando como base a Nota Técnica "Perfil das Escolas Similares" 2014 do INEP. **Metodologia:** Empregou-se a análise de redes sociais e a modelagem a partir de dados abertos governamentais para investigar a interconexão entre as escolas, considerando suas características semelhantes e indicadores educacionais. **Resultados:** Os dados abertos englobam informações sobre desempenho escolar e níveis socioeconômicos, processados e analisados por meio de algoritmos e do desenvolvimento de um software específico. O estudo resultou na visualização web de uma rede multidimensional, com diversas possibilidades de análise, evidenciando que, das 408 escolas, cada uma estabelece conexões com pelo menos 70 outras escolas em média. **Conclusão:** Essa interconexão sugere a tendência das escolas em compartilhar características semelhantes, proporcionando potencial para promover colaboração e compartilhamento de recursos e práticas educacionais bem-sucedidas, contribuindo para uma melhor governança na educação na região.

Palavras-chave: Dados abertos governamentais; Governança educacional; Visualização da informação; Análise de redes complexas; Rede municipal de ensino – Salvador/Bahia.

Abstract

Introduction: The objective of this study is to represent similar schools located in the urban area of the Microregion of Salvador - Bahia, in order to identify possible paths for collaboration among them, using as a basis the Technical Note "Profile of Similar Schools" 2014 of INEP. Methodology: We employed social network analysis and modeling from open government data to investigate the interconnection between schools, considering their similar characteristics and educational indicators. Results: The open data encompass information on school performance and socioeconomic levels, processed and analyzed through algorithms and the development of a specific software. The study resulted in the web visualization of a multidimensional network, with several possibilities for analysis, showing that, of the 408 schools, each establishes connections with at least 70 other schools on average. Conclusion: This interconnection suggests the tendency of schools to share similar characteristics, providing potential to promote collaboration and sharing of resources and successful educational practices, which contributes to better governance in education in the region.

Keywords: Government open data; Educational governance; Information visualization; Complex network analysis; Municipal education network - Salvador/Bahia

Resumen

Introducción: El objetivo de este estudio es representar las escuelas similares ubicadas en el área urbana de la Microrregión de Salvador - Bahía, con el fin de identificar posibles caminos de colaboración entre ellas, utilizando como base la Nota Técnica "Perfil de Escuelas Similares" 2014 del INEP. Metodología: Se empleó análisis de redes sociales y modelización a partir de datos abiertos de gobierno para investigar la interconexión entre escuelas, considerando sus características similares e indicadores educativos. Resultados: Los datos abiertos abarcan información sobre desempeño escolar y niveles socioeconómicos, procesada y analizada a través de algoritmos y el desarrollo de un

software específico. El estudio resultó en la visualización web de una red multidimensional, con varias posibilidades de análisis, mostrando que, de las 408 escuelas, cada una establece conexiones con al menos otras 70 escuelas en media. Conclusión: Esta interconexión sugiere la tendencia de las escuelas a compartir características similares, proporcionando potencial para promover la colaboración y el intercambio de recursos y prácticas educativas exitosas, lo que contribuye a una mejor gobernanza en la educación en la región.

Palabras clave: Datos abiertos gubernamentales; Gobernanza educativa; Visualización de información; Análisis de redes complejas; Red municipal de educación - Salvador/Bahia.

1. Introdução

Nos últimos anos, o estudo de dados abertos associados aos estudos de redes complexas tem se mostrado uma ferramenta promissora para a compreensão de diversos fenômenos em várias áreas do conhecimento. No campo da educação, a análise dos dados governamentais abertos através de redes pode ser útil para identificar padrões de similaridade entre escolas, a fim de orientar políticas públicas e melhorar a qualidade do ensino, além de possibilitar a criação de estratégias e estimular a colaboração e a difusão do conhecimento científico e social (CARNEIRO *et al.*, 2018).

O objetivo deste estudo é apresentar um modelo de construção do conhecimento baseado em redes complexas, alimentado por dados abertos governamentais, para identificar as similaridades entre as escolas. A motivação para essa iniciativa surge a partir da Nota Técnica intitulada “Perfil Escolas Similares”, emitida no ano de 2014 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Essa nota técnica aborda escolas com características semelhantes, bem como as limitações em sistemas e modelos de processamento e visualização de redes com dados abertos escolares.

2. Procedimentos Metodológicos

Trata-se de uma pesquisa exploratória, com abordagem quali-quantitativa, utilizando a análise de redes, para possibilitar a visualização e descoberta de relações do objeto de análise a outros fenômenos. Para representação dos dados de uma pesquisa com um excesso de dados e informações, o uso de critérios e indicadores auxiliará também na interpretação, com base nas fundamentações teóricas diante da análise da rede criada.

2.1 Classificação do perfil das escolas similares

A normativa “Perfil das Escolas Similares” orienta que as escolas estejam organizadas a partir de três critérios: a) Microrregião geográfica; b) Tipo de localização (Rural ou Urbana); c) e aproximação entre os valores absolutos do Indicador de Nível Socioeconômico (Inse) (BRASIL, 2014).

Para realizar a análise da normativa, utiliza-se uma metodologia baseada na classificação das escolas para promover a análise de agrupamento (*cluster*). A análise de agrupamento proposta, pretende identificar escolas com características comuns que possam estar relacionadas ao seu desempenho, permitindo que ações específicas sejam tomadas para melhorar seu desempenho. Segundo o INEP, “[...] os agrupamentos de escolas permitem a identificação de padrões de desempenho que podem ser úteis na elaboração de políticas e ações mais efetivas, já que elas são dirigidas a grupos de escolas com características e necessidades comuns” (BRASIL, 2014, p. 2).

A classificação é orientada pela construção de dois grandes grupos guiados pela localização (Rural ou Urbana); em seguida, a estruturação dos subgrupos, aglutinados em número máximo de 16 escolas. Após essa primeira classificação, a normativa propõe dois métodos para o cálculo do perfil das escolas similares:

- O cálculo consiste na soma da quantidade de alunos em cada nível da escala de proficiência de língua portuguesa e matemática.
- O cálculo consiste na média ponderada, composta pela quantidade de alunos, das médias de proficiências de língua portuguesa e matemática das escolas pertencentes ao respectivo Subgrupo.

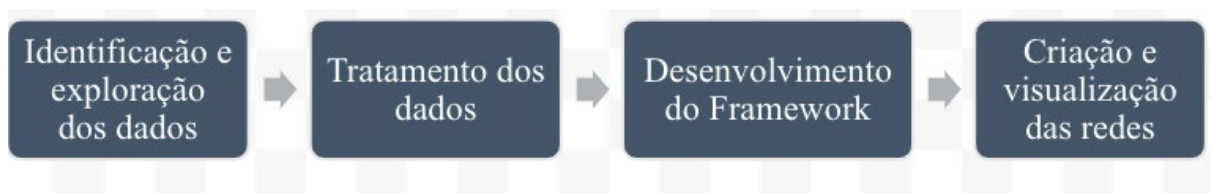
A normativa afirma que ambos os métodos permitem “[...] sintetizar o resultado, para cada disciplina ou componente curricular, de escolas com características semelhantes em termos socioespaciais e sociodemográficos, possibilitando a análise de maneira contextualizada” (BRASIL, 2014, p. 5). No entanto, não especifica caminhos de visualização e/ou demais instruções para análises dos resultados obtidos pelos métodos.

Deste modo, em busca de ampliar as possibilidades de análise e de relações entre esses grupos de escolas, optamos pela modelagem de redes.

2.2 Modelagem de redes de escolas similares

O estudo das redes complexas é um importante caminho científico para representação dos fenômenos sociais e do conhecimento (Carneiro; PEREIRA; GALEFFI, 2019), e este foi o caminho seguido no processo de modelagem deste estudo, desenvolvido em quatro etapas:

Figura 1 – Etapas das modelagens das redes de escolas similares.



Fonte: Elaboração própria (2023).

Utilizamos, na modelagem os conjuntos de dados abertos compostos pelas planilhas dos Resultados do Índice de Desenvolvimento da Educação Básica (Ideb) 2019 por escola; associadas às planilhas de Taxa de Rendimento Escolar para o ano 2019 por escola; e a planilha da distribuição do Inse por escola para o ano 2019. Após a exportação das planilhas, estabelecemos os descritores e metadados para a identificação dos dados.

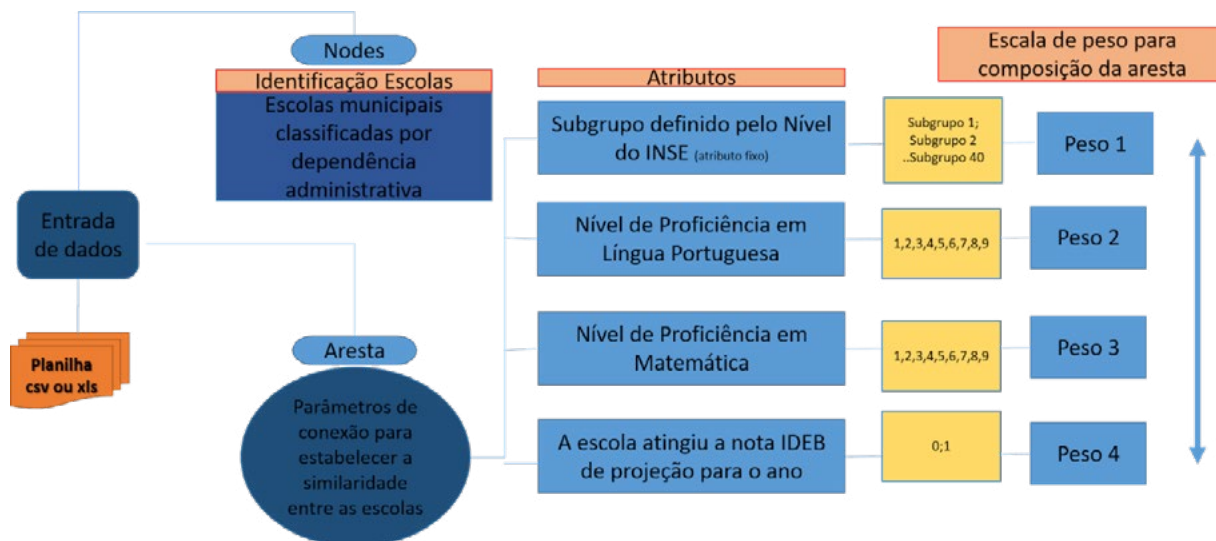
Na etapa da exploração dos dados, a amostra para este estudo foi definida pela zona urbana da Microrregião de Salvador – Bahia – Nordeste – Brasil. A microrregião é composta por 10 municípios, sendo eles: Camaçari, Candeias, Dias D'Ávila, Itaparica, Lauro de Freitas, Madre de Deus, Salvador, São Francisco do Conde, Simões Filho e Vera Cruz (BRASIL, 2022).

Em seguida, na fase de tratamento de dados, foram feitas as exclusões das escolas sem dados no período, sem identificação de localidade, duplicadas ou com código escolar divergente da base do INEP. O tratamento dos dados resultou em uma amostra de 408 escolas, com 21.681 estudantes matriculados.

Após a consolidação da amostra, os dados foram classificados seguindo as recomendações da Norma Técnica: Perfil das Escolas Similares (BRASIL, 2014), dando origem a 26 subgrupos que representam as escolas municipais da microrregião de Salvador – Bahia, localizadas na zona urbana das cidades.

Para o desenvolvimento das relações que deram origem à rede, fez-se necessário a construção de um *framework* denominado *SchoolSimilar 1.0* (Figura 2).

Figura 2 – Modelagem do framework SchoolSmiliar 1.0.



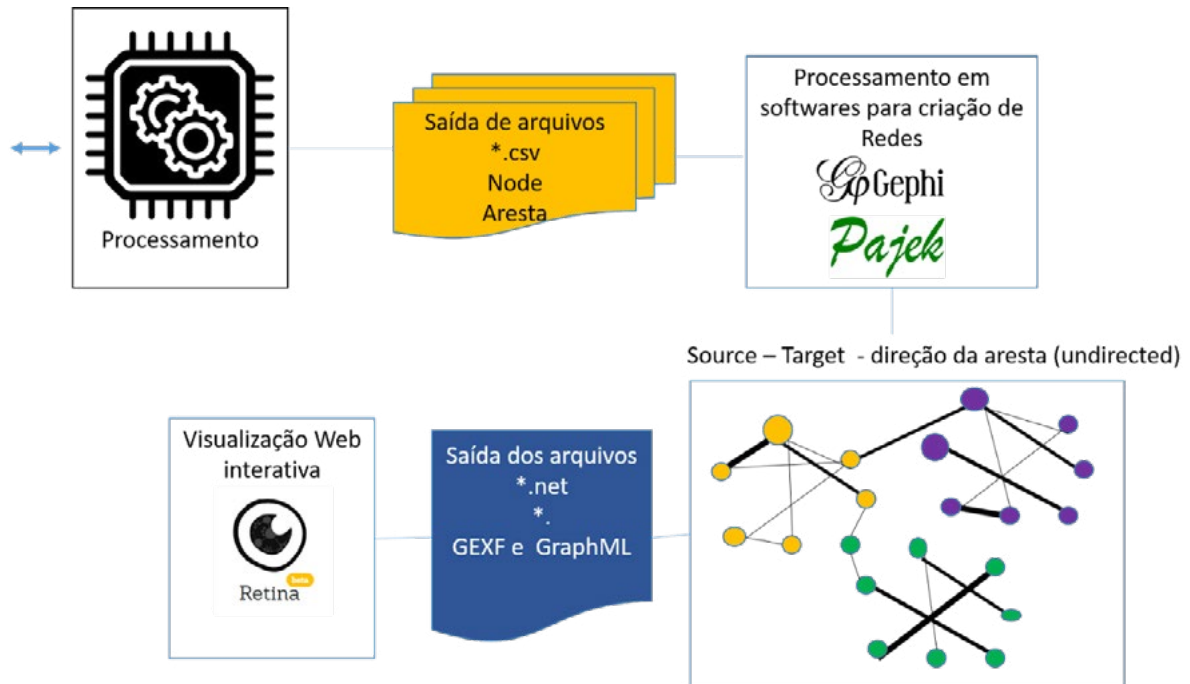
Fonte: Elaboração própria (2023).

No *SchoolSimilar 1.0*, apresentado na Figura 2, são definidas as entidades da rede através da utilização do Código da escola como identificação para a escola, que passam a representar os Vértices <V>; e os atributos de: a) Subgrupo; b) Nível de proficiência em língua portuguesa; c) Nível de proficiência em Matemática; d) Projeção atingida. Esses são os atributos de relacionamento que irão compor a construção das Arestas <E>.

O *framework* atua em três ciclos para a composição das arestas (Figura 2): (1) estabelece a conexão entre as escolas que compõem o mesmo subgrupo; (2) estabelece a conexão entre as escolas que não pertencem ao mesmo subgrupo, mas estão enquadradas no mesmo nível de proficiência em Língua Portuguesa e/ou Matemática, e entre as escolas que atingiram a nota IDEB correspondente à projeção para aquele ano; e por fim, (3) atribui peso à aresta conforme o número de atributos similares entre as escolas. Após essas três etapas, inicia-se a fase de criação e visualização do grafo de uma rede não dirigida que representa o perfil das escolas similares.

A sequência ilustrada na Figura 3, mostra o processamento computacional que gera os arquivos de saída em diferentes formatos, como **csv*, **graphml* e **gexf*. Para manipular e analisar as redes, utilizamos os *softwares Gephi 10.0.1* e *Pajek 5.3*. Esses programas também foram empregados para aplicar algoritmos estatísticos que auxiliam na obtenção de informações matemáticas relevantes das redes analisadas.

Figura 3 – Criação e visualização da rede.



Fonte: Elaboração própria (2023).

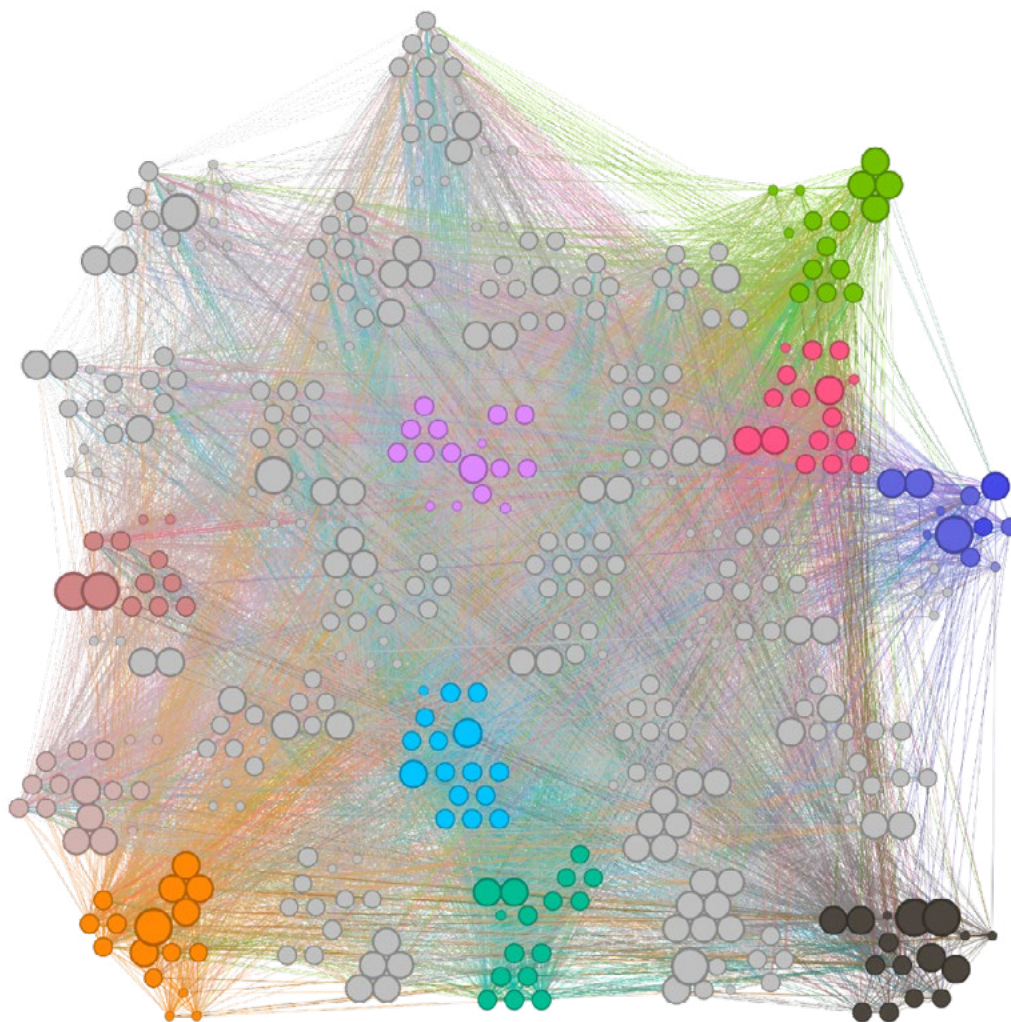
Nesta etapa de visualização, apresenta-se uma modelagem que busca envolver o usuário por meio da visualização interativa proposta pelo projeto Retina¹. Com a visualização da rede na *web*, o usuário tem a possibilidade de aplicar filtros de acordo com seus interesses, permitindo uma análise mais aprofundada dos perfis das escolas.

1 [Retina](#) é construído com *OuestWare*, *CNRS CIS* e *Tommaso Venturini*

3. Discussão e Resultados

Apresenta-se a análise dos dados e a estrutura empregada para examinar a rede (Figura 4). Os índices selecionados para descrever e investigar a rede² incluem: número de vértices ($n = |V|$), número de arestas ($m = |E|$), grau médio ($\langle k \rangle$), diâmetro (D), coeficiente médio de aglomeração (C), caminho médio mínimo (L) e densidade (Δ).

Figura 4 - Rede de Escolas Similares da Microrregião de Salvador – Bahia – Zona urbana (2023).



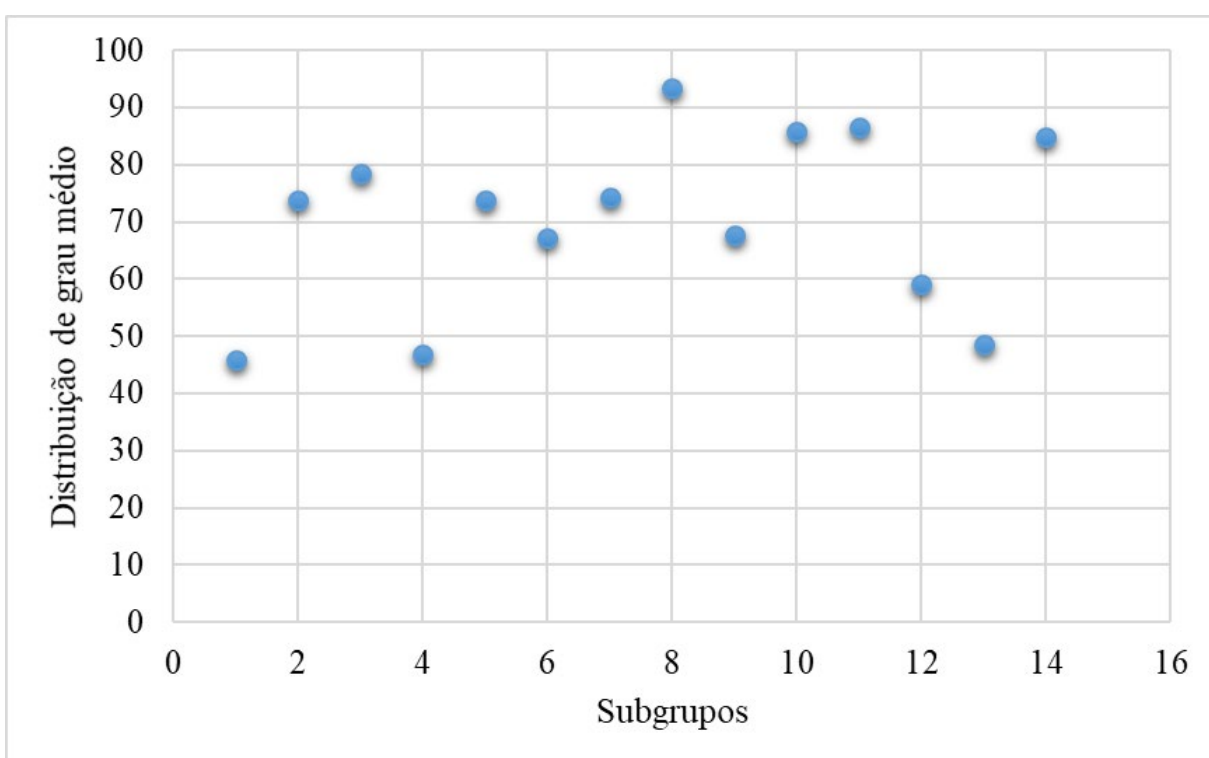
Fonte: Elaboração própria (2023).

² A visualização *web* da rede está disponível em: l1nq.com/similarschool.

A Rede das Escolas Similares situadas na área urbana da Microrregião de Salvador, na Bahia, é formada por 408 nós e 14.881 conexões. Em média, cada instituição de ensino estabelece vínculos com aproximadamente 70 outras escolas, demonstrando um alto grau de interconexão entre elas. Essa tendência de conexão entre as escolas indica que compartilham características comuns, o que pode facilitar a colaboração e o intercâmbio de recursos e práticas educacionais bem-sucedidas na região.

Ao examinar a distribuição dos graus de cada subgrupo que compõem a rede, é possível observar na Figura 5, a quantidade de conexões estabelecidas além dos próprios grupos para os 14 primeiros subgrupos.

Figura 5 - Distribuição de grau médio dos Subgrupos 1 a 14.



Fonte: Elaboração própria (2023).

A análise da distribuição do grau médio nos subgrupos 1 a 14, conforme ilustrado na Figura 5, permite identificar que as escolas com o perfil mais notável possuem proficiência de nível 4 em Língua Portuguesa e Matemática e atingiram a meta do IDEB projetada para 2019.

Examinando outras medidas da rede, notamos que a maior distância entre quaisquer escolas é representada pelo Diâmetro (D) de 3, indicando que a distância máxima entre duas escolas é de apenas 3 etapas. O coeficiente médio de aglomeração (C), que mede a tendência dos nós se agruparem, é de 0,694, sugerindo que as escolas estão bastante agrupadas em clusters. O caminho médio mínimo (L), que representa a média de todas as distâncias mais curtas entre quaisquer dois nós, é de 1,88, demonstrando que, em média, menos de 2 escolas precisam ser percorridas para se conectar de uma escola a outra na rede. A densidade (Δ) de 0,174 indica que menos de um quinto de todas as conexões possíveis entre as escolas existem na rede.

Essas medidas auxiliam na compreensão da estrutura global da rede de similaridade entre escolas, abordando a distância máxima entre elas, a tendência de agrupamento em clusters e a eficiência geral da rede em conectar as instituições de ensino.

4. Considerações Finais

Este estudo propôs um modelo de análise de dados abertos governamentais visando identificar similaridades entre escolas, para promover a detecção de instituições com maior potencial para colaboração mútua, baseada em suas semelhanças. Esse modelo tem caráter inovador, ao favorecer o estímulo no desenvolvimento de processos colaborativos por meio do uso de dados educacionais abertos.

Adicionalmente, a aplicação de metodologias fundamentadas em redes complexas no âmbito educacional pode colaborar na percepção de tendências e desafios compartilhados por escolas similares.

Assim, a cooperação entre escolas pode ser fomentada, estabelecendo um ambiente onde a troca de conhecimentos e experiências entre profissionais da educação enriqueça o processo de ensino-aprendizagem e, por consequência, gere impactos positivos no desempenho dos estudantes.

Referências

BRASIL. Instituto Brasileiro de Geografia e Estatística (IBGE). **Microrregiões**. 2022. Disponível em: <https://bit.ly/3IH-cmcB>. Acesso em: 5 abr. 2023.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Nota técnica**: perfil escolas similares. Brasília: INEP, 2014. Disponível em: <https://bit.ly/3WDArXl>. Acesso em: 1 abr. 2023

CARNEIRO, Tereza Kelley Gomes; BARRETO, Renata Souza Freitas Dantas; MONTEIRO, Roberto Luiz Souza; PEREIRA, Hernane Borges de Barros. A distance learning pedagogical management model based on affinity networks. **Obra Digital**: journal of communication and technology, Barcelona, v. 14. p. 13-33, 2008. Disponível em: <http://revistesdigitals.uvic.cat/index.php/obradigital/article/view/146>. Acesso em: 1 abr. 2023

CARNEIRO, Tereza Kelly Gomes; PERREIRA, Hernane Borges de Barros; GALEFFI, Dante Augusto. **Redes de afinidade como estratégia de gestão pedagógica e difusão do conhecimento em cursos na modalidade a distância**. Construção e difusão do conhecimento: conectando saberes. Salvador: EDIFBA, 2019.