

NOVO MÉTODO PARA REPRESENTAÇÃO DE FENÔMENOS SOCIAIS MULTIDIMENSIONAIS: SMART K-MEANS

Daniel Cirino Martins

Universidade Estadual de Montes Claros
Campus Universitário Prof. Darcy Ribeiro, Av. Prof. Rui Braga, s/n - Vila Mauriceia, Montes
Claros - MG, 39401-089
it.danielcirino@gmail.com

Marcos Flávio Silveira Vasconcelos D'Angelo

Universidade Estadual de Montes Claros
marcos.dangelo@unimontes.br

Matheus Pereira Libório

Universidade Estadual de Montes Claros
m4th32s@gmail.com

Ekel, Petr

Pontifícia Universidade Católica de Minas Gerais
petr.ekel2709@gmail.com

Honovan Paz Rocha

Universidade Estadual de Montes Claros
honovan.rocha@ufvjm.edu.br

Douglas Alexandre Gomes Vieira

Centro Federal de Educação Tecnológica de Minas Gerais
douglas.agv@gmail.com

Laura Cozzi Ribeiro

Pontifícia Universidade Católica de Minas Gerais
laura.cozzi.ribeiro@gmail.com

Maria Fernanda Oliveira Guimarães

Pontifícia Universidade Católica de Minas Gerais
mariafernanda.guimaraes810@gmail.com

RESUMO

Esta pesquisa exploratória propõe um novo método para representação de fenômenos multidimensionais baseado no algoritmo k-means. O denominado “smart k-means” se utiliza do índice de entropia de Shannon para identificar e descartar indicadores individuais com baixa diversidade informacional, que não distinguem as unidades de tomada de decisão adequadamente. Esse processo de descarte garante grupos com coesão e resolução aceitáveis, e dentro do número de arranjos de grupos desejado pelo tomador de decisão. Os resultados da aplicação do novo método na representação da exclusão social de Maringá mostram que o processo de descarte de indicadores individuais de baixa diversidade informacional assegura a obtenção de grupos internamente homogêneos e externamente heterogêneos. Esses resultados contribuem para uma representação confiável do fenômeno multidimensional, facilitando a sua compreensão e a elaboração de políticas públicas para sua redução.

PALAVRAS CHAVE. Problemas multidimensionais. Indicadores compostos. Diversidade informacional.

2. ADM – Apoio à Decisão Multicritério / 5.BDA – Big Data e Analytics

NEW METHOD FOR REPRESENTING MULTIDIMENSIONAL SOCIAL PHENOMENA: SMART K-MEANS

ABSTRACT

This exploratory research proposes a new method for representing multidimensional phenomena based on the k-means algorithm. The so-called “smart k-means” use the Shannon entropy index to identify and discard individual indicators with low informational diversity, which do not distinguish decision-making units adequately. This discard process ensures groups have acceptable cohesion and resolution and are within the number of group arrangements desired by the decision maker. The results of applying the new method to represent social exclusion in Maringá show that discarding individual indicators of low informational diversity ensures obtaining internally homogeneous and externally heterogeneous groups. These results contribute to a reliable representation of the multidimensional phenomenon, facilitating its understanding and the development of public policies to reduce it.

KEYWORDS. Multidimensional problems. Composite indicators. Informational diversity.

2. ADM – Apoio à Decisão Multicritério / 5.BDA – Big Data e Analytics

1. Introdução

O algoritmo k-means tem sido amplamente utilizado para representar fenômenos sociais multidimensionais e auxiliar a elaboração de políticas públicas para redução de desigualdades [Chamboko et al. 2017], [Basel et al. 2021], [Kallingal e Mohammed Firoz 2023]. Em particular, o método permite classificar as diferentes áreas das cidades com base nos múltiplos indicadores individuais (renda, nível de educação, condições dos domicílios), facilitando a compreensão dos diversos elementos presentes no fenômeno multidimensional [Libório et al. 2022].

Embora extremamente úteis para construção de indicadores compostos sociais, o k-means apresenta limitações importantes, que podem levar a interpretações equivocadas e políticas públicas ineficazes [European Union and Joint Research Center 2008]. Entre essas limitações está a definição de agrupamentos de baixa confiabilidade e interpretabilidade [Bernardes et al. 2022]. Isso ocorre porque o número de agrupamentos deve ser definido a priori, e porque técnicas de definição de número ideal de grupos geram resultados contraditórios e que não garantem a obtenção de grupos com coesão e resolução aceitáveis [Kassambara 2017], [Yuan e Yang 2019].

Esta pesquisa exploratória propõe um novo método que alia confiabilidade e interpretabilidade na definição do número de agrupamentos, para a adequada representação de fenômenos sociais multidimensionais. O denominado “smart k-means” encontra, dentro do número de arranjos de grupos desejado pelo tomador de decisão, aquele que apresentam coesão e resolução, dentro dos limiares de aceitação, preservando o máximo dos dados de entrada através de um processo iterativo de exclusão de indicadores individuais com alta incerteza informacional.

As vantagens da aplicação do método são demonstradas através do caso da exclusão social da cidade de Maringá, Paraná. Os resultados mostram um ganho significativo em termos de preservação dos dados de entrada, garantia da confiabilidade e interpretabilidade dos resultados. Essas vantagens tornam o método útil para sociólogos, geógrafos e demais pesquisadores das áreas sociais, bem como contribui para a formulação de políticas públicas de redução de desigualdades. Além disso, as vantagens do método são extensíveis a outros tipos de representação de fenômenos de natureza multidimensional nas áreas ambiental, econômica e de negócios.

2. Revisão de Literatura

A exclusão social é um fenômeno de natureza multidimensional, que não pode ser medido através de um indicador individual como o rendimento médio ou o nível de escolaridade [van Bergen et al. 2014], [Keogh et al. 2021], [Jiang et al. 2022]. Assim, medir a exclusão social requer a operacionalização de um framework que permita a construção dos chamados indicadores compostos [European Union and Joint Research Center 2008]. O framework operacional dos indicadores compostos envolve um conjunto de métodos e técnicas para a consideração simultânea de múltiplos indicadores individuais associados ao fenômeno multidimensional em uma medida de síntese [Abello et al. 2016], [García-Vélez et al. 2020], [Vela-Jiménez e Sianes 2021].

A representação da exclusão social usada nesta pesquisa é baseada no framework conceitual proposto no livro “Desigualdades intraurbanas metodologias para produção e análise de indicadores compostos” [Martinuci e Libório 2022]. Esse framework organiza a exclusão social entre as cinco dimensões e quinze indicadores individuais mostradas na Tabela 1.

Tabela 1. Dimensões e indicadores individuais de exclusão social usados na pesquisa.

Dimensão	Subindicador	Descrição
Condições dos domicílios	HS1	Percentual de domicílios sem banheiro;
	HS2	Percentual de domicílios com mais de 3 banheiros;
	HS3	Percentual de domicílios com rede de esgoto;
	HS4	Percentual de domicílios com rede de água;
	HS5	Percentual de domicílios alugados ou cedidos.
Demografia	DM1	Número de habitantes por domicílio;
	DM2	Percentual de chefes de família entre 10-19 anos;
	DM3	Razão de dependência: crianças por chefe de família

	DM4	Número de crianças de até um ano de idade.
	EC1	Chefes de família sem rendimento;
Economia	EC2	Chefes de família com renda maior que 20 salários-mínimos;
	EC3	Chefes de família com renda de até 02 salários-mínimos.
Educação	ED1	Percentual de pessoas 10-14 anos não alfabetizadas;
	ED2	Chefes de família não alfabetizados.
Meio ambiente	EN1	Arborização, cobertura vegetal

A síntese dos indicadores individuais de exclusão social em indicadores compostos pode ser construída através de duas categorias de métodos. Os métodos multicritério [El Gibari et al. 2019] têm a particularidade de ponderação exógena de indicadores individuais e incluem o Analytic Hierarchy Process (AHP) [Sutadian et al. 2017], Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [Liern e Pérez-Gladish 2021], Multi-Attribute Value Theory (MAVT) [Figueira et al. 2023]. Por sua vez, os métodos estatísticos de construção de indicadores compostos [Fernandez & Martos 2020] definem os pesos dos indicadores individuais de forma endógena, utilizando os próprios dados, sendo o Benefit-of-the-Doubt [Karagiannis e Karagiannis 2018], Principal Component Analysis [Li et al. 2012] e Data Envelopment Analysis [Pereira et al. 2021] os mais utilizados.

Esses diversos métodos fornecem uma ampla gama de alternativas para lidar com problemas que ocorrem durante as etapas de normalização, ponderação e agregação dos indicadores individuais do indicador composto [Mazziotta e Pareto 2016], [Cinelli et al. 2022]. No entanto, novos métodos são continuamente introduzidos para superar problemas dos métodos tradicionais, ou para resolver novos problemas, como a desconsideração da heterogeneidade espacial [Fusco et al. 2018; 2020], [Postiglione et al. 2023], a incomparabilidade entre indicadores compostos ao longo do tempo [Mazziotta e Pareto 2022], e a ponderação baseada em dados de indicadores individuais com pesos incompatíveis com o referencial teórico do fenômeno multidimensional [Correa Machado et al. 2023], [Libório et al. 2024].

Embora ofereçam representações unidimensionais para uma ampla gama de fenômenos multidimensionais, esses métodos não classificam as unidades de tomada de decisão. A ausência dessa classificação é problemática para representações geográficas de muitas unidades de tomada de decisão, como no caso da exclusão social nas cidades. Por exemplo, a cidade de Maringá, de 409.657 mil habitantes, possui 483 unidades de tomada de decisão (setores censitários). Nesses casos, a análise e formulação de políticas a partir das classes de exclusão social das unidades de decisão é mais fácil e intuitiva do que a análise das pontuações de todas as unidades de decisão. A Figura 1 ilustra essa situação, tomando como base o indicador composto de exclusão social construído com os dados da Tabela 1.

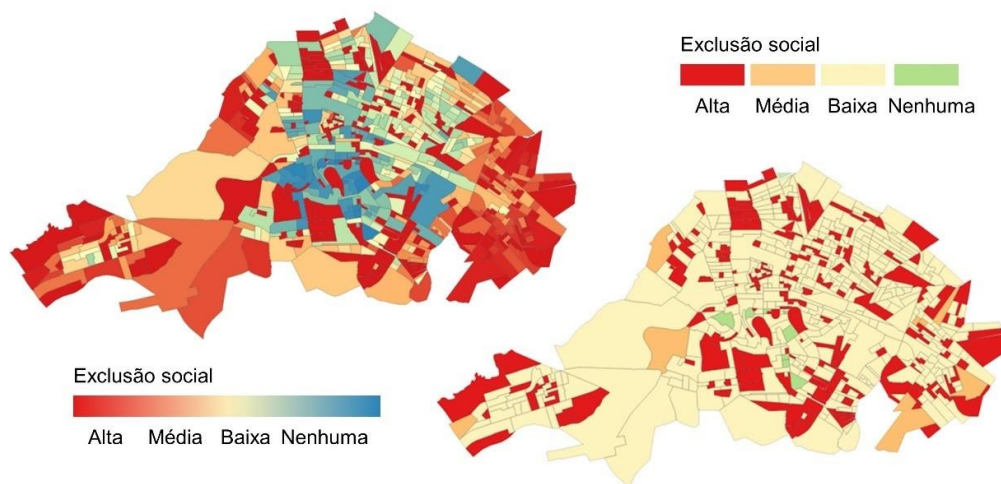


Figura 1. Indicador composto de exclusão social por pontuações e por ranking.

No framework operacional de indicadores compostos, o algoritmo de agrupamentos k-means [MacQueen 1967] é a solução natural para classificar a exclusão social de unidades de tomada de decisão com base na similaridade de seus indicadores individuais [European Union and Joint Research Center 2008].

A possibilidade de classificar os fenômenos em classes, para sua representação espacial torna o k-means um método de construção de indicadores compostos bastante popular [Chamboko et al. 2017], [Basel et al. 2021], [Kallingal e Mohammed Firoz 2023]. Entretanto, o k-means apresenta limitações importantes, que se repercutem na confiabilidade e interpretabilidade dos resultados.

A confiabilidade dos resultados, representada pela coesão (homogeneidade intragrupo) e resolução (heterogeneidade entre grupos), é determinada a *posteriori*, após a obtenção dos agrupamentos [Kassambara 2017], [Yuan e Yang 2019], [Libório et al. 2022]. Embora vários métodos permitam verificar o número ideal de agrupamentos a priori, esses métodos apresentam resultados contraditórios [Sai Krishna et al. 2018]. Além disso, métodos de definição de número de ideal de grupos podem apontar para soluções de difícil interpretabilidade, ou seja, quando o número ideal de grupos é elevado.

Para demonstrar essas duas limitações, foram calculados o número ideal de grupos para os dados de exclusão social de Maringá da Tabela 2, usando cinco métodos. Os resultados desses cálculos são apresentados na Figura 2, que revela números ideal de grupo diferentes para Gap statistic (14 grupos), Elbow (8 grupos) e Galinsk, Davies e Silhueta (2 grupos).

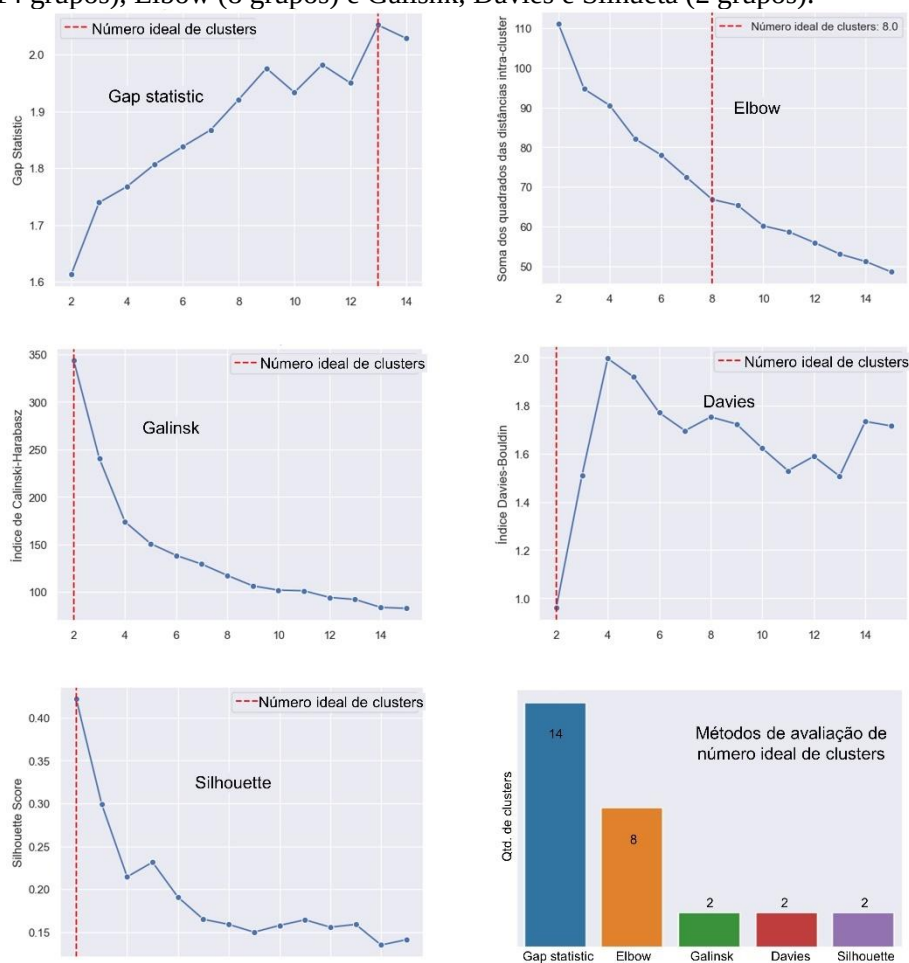


Figura 2. Definição do número ideal de grupos por diferentes métodos.

Além de oferecer resultados contraditórios e de difícil interpretabilidade (p.ex., Gap statistic e Elbow), métodos de definição de número ideal de grupos não garantem que a coesão

entre as unidades de tomada de decisão intragrupo e a resolução entre grupos é suficientemente confiável.

A Figura 3 mostra que nenhum dos agrupamentos sugeridos pelos testes anteriores apresentam qualidade adequada. Portanto, a quantidade sugerida pelos testes mais usados na literatura não permite agrupar os setores censitários da cidade de Maringá de maneira confiável. Todos os resultados apresentam silhueta média abaixo do limiar de 0,50 [Kassambara 2017]. Além disso, muitos setores censitários apresentam silhueta negativa nos arranjos com 8 e 14 grupos, indicando que foram posicionados em grupos errados [Libório et al. 2022].

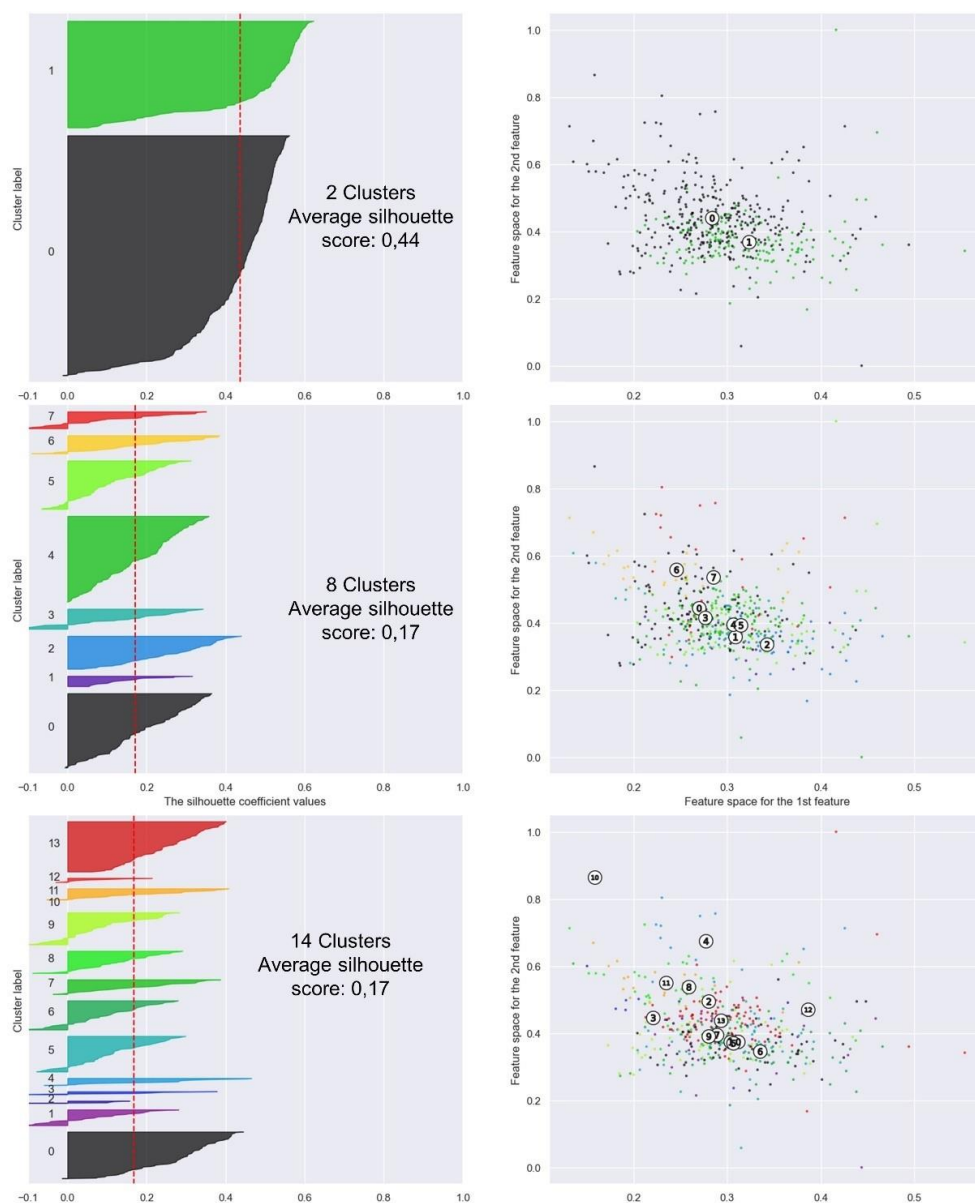


Figura 3. Gráfico de silhueta (coesão e resolução) para arranjos com 2, 8 e 14 grupos e visualização dos dados dos grupos.

Em suma, os gráficos da Figura 3 comprovam que nenhum dos arranjos de grupos apresenta oferece uma boa separação da exclusão social entre os setores censitários de Maringá. Essa limitação motiva o desenvolvimento do smart k-means, que ofereça agrupamentos com qualidade aceitáveis, para a adequada representação de fenômenos multidimensionais, como a exclusão social de Maringá.

3. Solução proposta: smart k-Means

O algoritmo “smart k-means” concilia confiabilidade e interpretabilidade na definição do melhor arranjo de grupos a ser utilizado na representação dos fenômenos sociais multidimensionais. O algoritmo busca dentro do intervalo definido pelo tomador de decisão, os arranjos de agrupamentos que alcançam coesão e resolução acima do limiar de aceitação.

Para isso, a pesquisa se baseia no princípio da diversidade informacional de Shannon [1948]. Com base nesse princípio, pesos maiores são atribuídos aos indicadores individuais com maior diversidade informacional [Karagiannis and Karagiannis 2018]. Em outras palavras, indicadores individuais com baixa diversidade informacional têm menos importância no indicador composto porque são menos capazes de diferenciar as unidades de tomada de decisão [Libório et al 2024].

O algoritmo do smart k-means ilustrado na Figura 4 segue essa lógica, descartando o subindicador do conjunto de dados com menor diversidade informacional até encontrar um arranjo com coesão e resolução com valores aceitáveis.

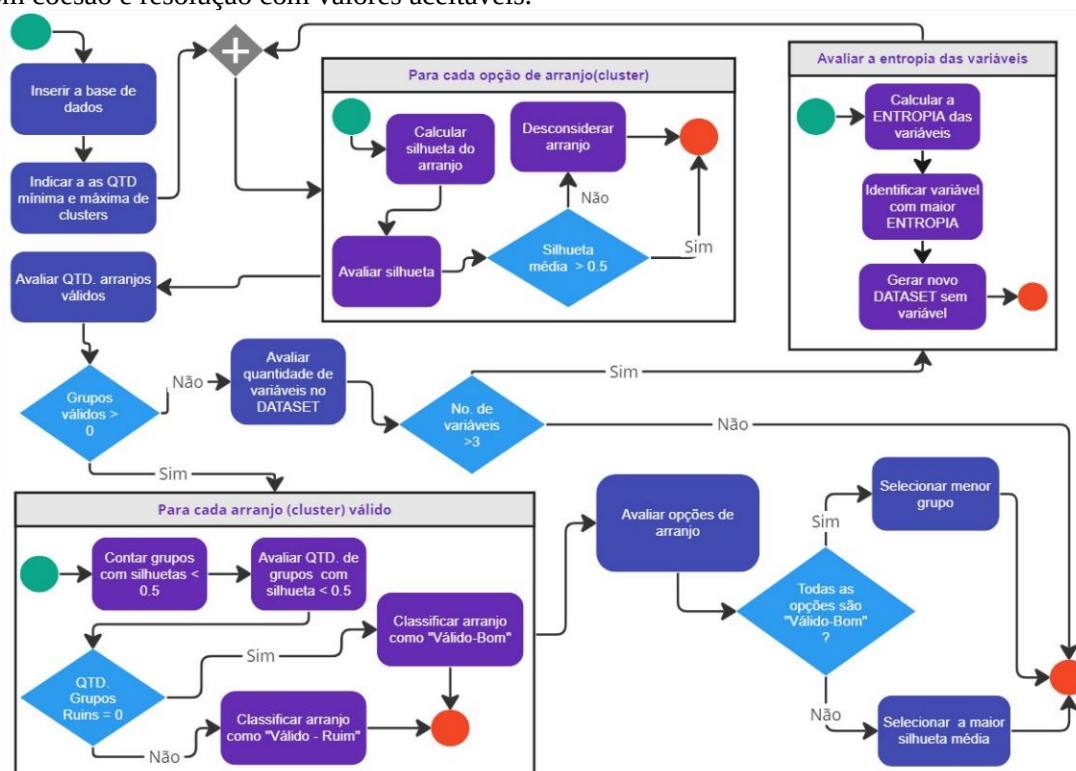


Figura 4. Algoritmo do “Smart k-means”.

Dentro da lógica do algoritmo, o cálculo do índice de entropia dos indicadores individuais é fundamental. Esse cálculo permite identificar a diversidade informacional dos indicadores individuais e sua ordem de descarte. A Tabela 2 mostra o índice de entropia para os indicadores individuais de exclusão social de Maringá.

Tabela 2. Índice de entropia (H) da informação dos indicadores individuais de exclusão social.

	EN1	EC3	HS5	DM3	EC1	ED2	HS3	DM4	HS2	DM2	DM1	EC2	HS4	ED1	HS1
H	8,92	8,82	8,75	8,69	8,57	8,27	8,16	7,04	6,80	6,66	6,42	4,35	4,24	3,58	2,09

Note que os indicadores individuais de maior diversidade informacional apresentam menor índice de entropia. Assim, a arborização, cobertura vegetal (EN1) e chefes de família com renda de até 02 salários-mínimos (EC3) são os dois primeiros indicadores individuais a serem descartados pelo algoritmo por terem maior índice de entropia, e menor diversidade informacional. Por sua vez, os indicadores individuais Percentual de pessoas 10-14 anos não alfabetizadas (ED1) e Percentual de domicílios sem banheiro (HS1) são os dois últimos a serem considerados no descarte devido a seu baixo índice de entropia, e alta diversidade informacional.

Neste ponto, cabe destacar o melhor desempenho do índice de entropia [Shannon, 1948] na exclusão de indicadores individuais. No exemplo da exclusão social de Maringá, essa abordagem encontra grupos com coesão e resolução aceitáveis após sete iterações, preservando oito indicadores individuais. Por sua vez, a abordagem baseada na exclusão de indicadores individuais com maior proporção de outliers e baixa carga fatorial manteve apenas seis indicadores individuais no indicador composto de exclusão social [Libório et al. 2022].

Além disso, o método permite que o tomador de decisão defina o intervalo de grupos que garanta a interpretabilidade do fenômeno multidimensional em análise. A Figura 5 mostra a qualidade do agrupamento obtido pelo “smart k-means” após a exclusão de 7 indicadores individuais. Neste exemplo, o intervalo de agrupamentos desejado pelo tomador de decisão foi estabelecido em 3 a 7 grupos.

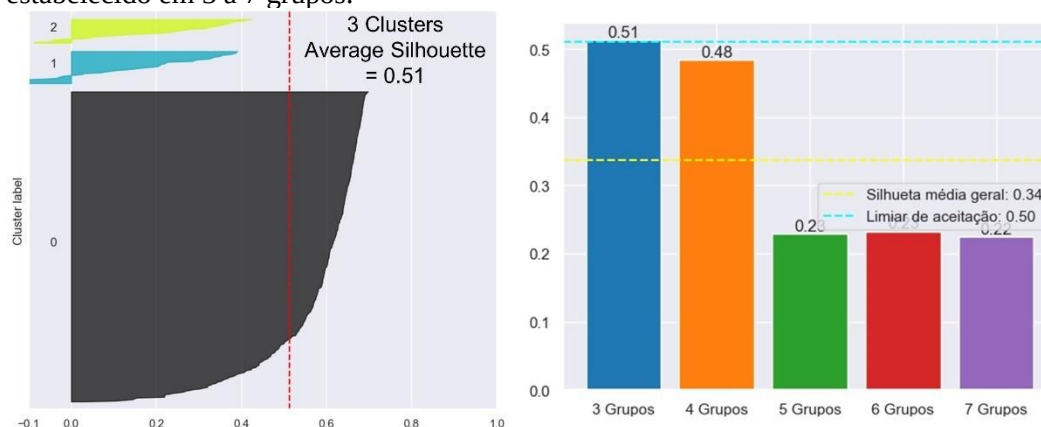


Figura 5. Solução encontrada pelo “smart k-means”.

Dentre dos possíveis arranjos definidos pelo tomador de decisão é possível observar claramente que a classificação da exclusão social em três grupos é confiável, pois apresenta silhueta média acima de 0,51. Esse limiar indica que os grupos de setores censitários estão bem separados, e que os setores censitários dentro de cada grupo são semelhantes. Além disso, é possível observar que os demais arranjos encontrados apresentam soluções com silhueta média inferior a de 3 grupos, indicando que a escolha do arranjo ideais é coerente.

Assim, pós o processamento do algoritmo, é possível representar a exclusão social de Maringá em três grupos, considerando os seguintes indicadores individuais:

- Número de habitantes por domicílio (DM1)
- Percentual de chefes de família entre 10-19 anos (DM2)
- Número de crianças de até um ano de idade (DM4)
- Chefes de família com renda maior que 20 salários-mínimos (EC2)
- Percentual de pessoas 10-14 anos não alfabetizadas (ED1)
- Percentual de domicílios sem banheiro (HS1)
- Percentual de domicílios com mais de 3 banheiros (HS2)
- Percentual de domicílios com rede de água (HS4)

Finalmente, é possível definir os grupos de exclusão social da cidade de Maringá com base nos valores médios dos indicadores individuais mantidos no modelo. Esses três grupos de exclusão social estão representados pelo mapa da Figura 6.

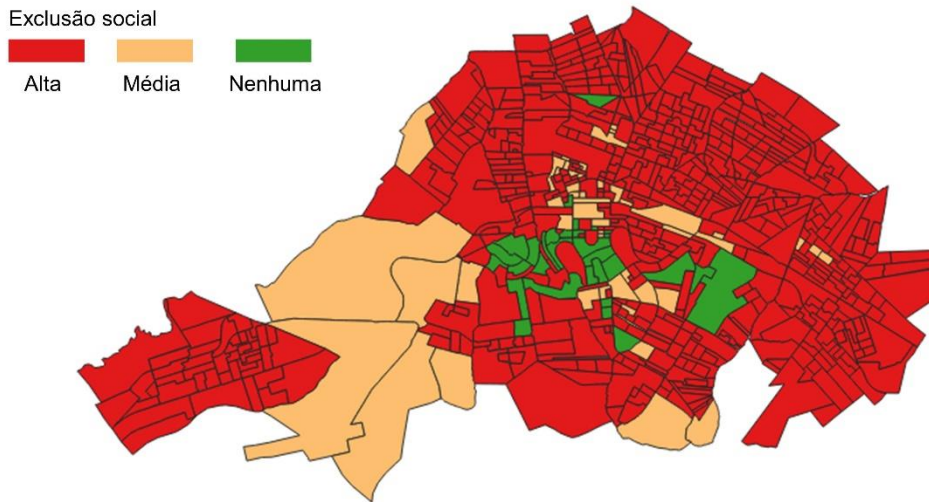


Figure 6. Mapa de exclusão social de Maringá.

O mapa mostra uma prevalência de setores censitários classificados como alta exclusão social. Por sua vez, é possível observar que os setores censitários classificados como média exclusão social estão concentrados na região meio oeste da cidade, em algumas áreas do centro e centro sul da cidade. Finalmente, é possível observar um menor número de setores censitários classificados como sem exclusão social. Esses setores censitários estão muito próximos e concentrados na área central da cidade.

4. Conclusões

Este estudo exploratório apresentou um novo método para representação de fenômenos sociais multidimensionais baseado no k-means. Importante destacar o aspecto inovador do método, que busca melhorar a coesão e resolução dos agrupamentos através da exclusão iterativa dos indicadores individuais com menor diversidade informacional. Em particular, o “smart k-means” se utiliza do índice de entropia de Shannon [Shannon, 1948] para identificar os indicadores individuais com menor diversidade informacional, e que não oferecem uma distinção adequada entre as unidades de tomada de decisão.

O método foi aplicado para representação do fenômeno multidimensional da exclusão social para cidade de Maringá, Paraná. Após 7 iterações, o algorítmico encontrou um arranjo de grupos com coesão e resolução acima do limiar de aceitação, garantindo a qualidade da representação da exclusão social.

Apesar do algoritmo oferecer uma solução confiável e de fácil interpretação, os resultados da pesquisa possuem limitações. Primeira, faltam simulações que garantam a aplicabilidade e robustez do algoritmo para outros dados de entrada. Segunda, faltam comparações que comprovem a vantagem de priorizar o descarte dos indicadores individuais pelo índice de entropia em detrimento do percentual de outliers, correlação média, ou da carga fatorial dos indicadores individuais. Terceira, o algoritmo não fornece informações sobre as características dos grupos, o que exige um esforço adicional do tomador de decisão para analisar o fenômeno multidimensional.

Agradecimentos

Esta pesquisa foi apoiada pela FAPEMIG (APQ-01042-23 e PCE-00437-24) e CNPq (407180/2023-0, 311922/2021-0 e 308265/2022-0).

5. Referências

- Abello, A., Cassells, R., Daly, A., D'Souza, G., & Miranti, R. (2016). Youth social exclusion in Australian communities: A new index. *Social Indicators Research*, 128, 635-660.
- Aparicio, J. and Kapelko, M., 2019. Enhancing the measurement of composite indicators of corporate social performance. *Social Indicators Research*, 144, pp.807-826.
- Aparicio, J., & Kapelko, M. (2019). Enhancing the measurement of composite indicators of corporate social performance. *Social Indicators Research*, 144, 807-826.
- Bernardes, P., Ekel, P. I., Rezende, S. F. L., Pereira Júnior, J. G., dos Santos, A. C. G., da Costa, M. A. R., .. & Libório, M. P. (2022). Cost of doing business index in Latin America. *Quality & Quantity*, 56(4), 2233-2252.
- Chamboko, R., Re, A., & Guvuriro, S. (2017). Mapping patterns of multiple deprivation in Namibia. *International Journal of Social Economics*, 44(12), 2486-2499.
- Correa Machado, A. M., Ekel, P. I., & Libório, M. P. (2023). Goal-based participatory weighting scheme: Balancing objectivity and subjectivity in the construction of composite indicators. *Quality & Quantity*, 57(5), 4387-4407.
- El Gibari, S., Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: a review. *Journal of Business Economics*, 89(1), 1-24.
- European Union and Joint Research Centre (2008). Handbook on constructing composite indicators: methodology and user guide. OECD publishing.
- Fernandez, E. J., & Martos, M. J. R. (2020). Review of some statistical methods for constructing composite indicators. *Studies of Applied Economics*, 38(1).
- Figueira, J. R., Oliveira, H. M., Serro, A. P., Colaço, R., Froes, F., Cordeiro, C. R., .. & Guimarães, M. (2023). A multiple criteria approach for building a pandemic impact assessment composite indicator: The case of COVID-19 in Portugal. *European journal of operational research*, 309(2), 795-818.
- Fusco, E., Vidoli, F., & Rogge, N. (2020). Spatial directional robust Benefit of the Doubt approach in presence of undesirable output: An application to Italian waste sector. *Omega*, 94, 102053.
- Fusco, E., Vidoli, F., & Sahoo, B. K. (2018). Spatial heterogeneity in composite indicator: A methodological proposal. *Omega*, 77, 1-14.
- García-Vélez, D. F., Contreras-Jaramillo, M. A., Torres-Gutiérrez, T. P., & Correa-Quezada, R. F. (2020). Social exclusion in Ecuador with housing indicators: A regional analysis. *Entrepreneurship and the community: A multidisciplinary perspective on creativity, social challenges, and business*, 159-176.
- Jiang, S., Jiang, C., Cheng, Y., & Li, W. (2022). Multidimensional measurement of child social exclusion: Development and psychometric properties of the social exclusion scale for children (SESC). *Children and Youth Services Review*, 141, 106624.

Jiménez-Fernández, E., Sánchez, A., & Ortega-Pérez, M. (2022). Dealing with weighting scheme in composite indicators: An unsupervised distance-machine learning proposal for quantitative data. *Socio-Economic Planning Sciences*, 83, 101339.

Kallingal, F. R., & Mohammed Firoz, C. (2023). Regional disparities in social development: A case of selected districts in Kerala, India. *GeoJournal*, 88(1), 161-188.

Karagiannis, R., & Karagiannis, G. (2018). Intra-and inter-group composite indicators using the BoD model. *Socio-Economic Planning Sciences*, 61, 44-51.

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1). Sthda.

Keogh, S., O'Neill, S., & Walsh, K. (2021). Composite measures for assessing multidimensional social exclusion in later life: Conceptual and methodological challenges. *Social Indicators Research*, 155, 389-410.

Li, T., Zhang, H., Yuan, C., Liu, Z., & Fan, C. (2012). A PCA-based method for construction of composite sustainability indicators. *The International Journal of Life Cycle Assessment*, 17, 593-603.

Libório, M. P., Karagiannis, R., Diniz, A. M. A., Ekel, P. I., Vieira, D. A. G., & Ribeiro, L. C. (2024). The Use of Information Entropy and Expert Opinion in Maximizing the Discriminating Power of Composite Indicators. *Entropy*, 26(2), 143.

Libório, M. P., Martinuci, O. D. S., Machado, A. M. C., Lyrio, R. D. M., & Bernardes, P. (2022). Time-Space Analysis of Multidimensional Phenomena: A Composite Indicator of Social Exclusion Through k-Means. *Social Indicators Research*, 159(2), 569-591.

Liern, V., & Pérez-Gladish, B. (2021). Building composite indicators with unweighted-TOPSIS. *IEEE Transactions on Engineering Management*, 70(5), 1871-1880.

MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

Martinuci, O. S. & Libório, M. P (2022). Desigualdades intraurbanas metodologias para produção e análise de indicadores compostos. Editora CRV, Curitiba.

Mazziotta, M., & Pareto, A. (2016, July). Methods for constructing non-compensatory composite indices: A comparative study. In *Forum for Social Economics* (Vol. 45, No. 2-3, pp. 213-229). Routledge.

Mazziotta, M., & Pareto, A. (2022). Normalization methods for spatio-temporal analysis of environmental performance: Revisiting the Min–Max method. *Environmetrics*, 33(5), e2730.

Pereira, M. A., Camanho, A. S., Figueira, J. R., & Marques, R. C. (2021). Incorporating preference information in a range directional composite indicator: the case of Portuguese public hospitals. *European Journal of Operational Research*, 294(2), 633-650.

Postiglione, P., Cartone, A., Andreano, M. S., & Benedetti, R. (2023). Constrained optimization for addressing spatial heterogeneity in principal component analysis: an application to composite indicators. *Statistical Methods & Applications*, 32(5), 1539-1561.

Sai Krishna, T. V., Yesu Babu, A., & Kiran Kumar, R. (2018). Determination of optimal clusters for a Non-hierarchical clustering paradigm K-Means algorithm. In *Proceedings of International Conference on Computational Intelligence and Data Engineering: ICCIDE 2017* (pp. 301-316). Springer Singapore.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.

Sutadian, A. D., Muttill, N., Yilmaz, A. G., & Perera, B. J. C. (2017). Using the Analytic Hierarchy Process to identify parameter weights for developing a water quality index. *Ecological indicators*, 75, 220-233.

van Bergen, A. P., Hoff, S. J., van Ameijden, E. J., & van Hemert, A. M. (2014). Measuring social exclusion in routine public health surveys: Construction of a multidimensional instrument. *PloS one*, 9(5), e98680.

Vela-Jiménez, R., & Sianes, A. (2021). Do current measures of social exclusion depict the multidimensional challenges of marginalized urban areas? Insights, gaps and future research. *International Journal of Environmental Research and Public Health*, 18(15), 7993.

Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *J*, 2(2), 226-235.