XXII ENGEMA
Encontro Internacional sobre Gestão Empresarial e Meio Ambiente

ISSN: 2359-1048
Novembro 2020

Análise de cluster para sistematização de indicadores de vulnerabilidade socioambiental

AMANDA ANSELMO DE MEDEIROS
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

ANA CECÍLIA FEITOSA DE VASCONCELOS
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

PAULO RIBEIRO LINS JUNIOR

Análise de cluster para sistematização de indicadores de vulnerabilidade socioambiental

Resumo:
A utilização dos indicadores tem sido basilar na formulação de políticas públicas, pois possibilitam uma análise atual, como também o monitoramento temporal, permitindo o acompanhamento das variáveis analisadas, a avaliação dos processos e planejamento de ações e, por conseguinte, a utilização adequada dos recursos investidos. Este artigo tem como objetivo sistematizar um conjunto de indicadores de vulnerabilidade socioambiental utilizando a análise de cluster baseada em similaridade textual. Para tanto, para cada uma das etapas propostas, foram indicadas técnicas matemáticas capazes de executar e viabilizar tais recomendações, implementadas através da linguagem de programação *Python.* Os resultados apresentam que os 98 indicadores foram agrupados em 46 clusters. A redução no quantitativo de variáveis e a identificação da similaridade entre os indicadores, proporcionam maior flexibilidade e abrangência da análise da vulnerabilidade socioambiental, viabilizando a análise das inter-relações existentes entre os indicadores e, por conseguinte, uma análise mais ampla e complexa.

Palavras-chave: Indicadores. Vulnerabilidade. Análise de Cluster. Análise de Similaridade.

1 INTRODUÇÃO

Indicadores podem ser compreendidos como uma medida que busca sintetizar informações relevantes de um dado fenômeno, analisando seu comportamento ao longo de um espaço temporal. Por serem percebidos como uma ferramenta importante no debate que envolve as questões de interesse público, quando bem elaborados, sua utilização pode proporcionar conclusões analíticas ou políticas de forma mais simples.

É neste contexto que se entende que para a elaboração de um sistema de indicadores, é necessário o embasamento teórico acerca do fenômeno estudado, o conhecimento de técnicas que viabilizem a sua execução, bem como sua contribuição para uma melhor compreensão do fenômeno que o sistema de indicadores envolve.

No caso específico desse estudo, que tem o fenômeno de vulnerabilidade socioambiental como cenário, é preciso compreender sua definição. Para Cutter (2011) vulnerabilidade corresponde ao "potencial para a perda". Para esta autora, a vulnerabilidade inclui tanto os "elementos de exposição ao risco" como os "fatores de propensão às circunstâncias que aumentam ou reduzem as capacidades da população, das infraestruturas ou dos sistemas físicos para responder e se recuperar de ameaças ambientais" (CUTTER; 2011, p.60).

É nesse sentido que os problemas advindos da poluição e degradação do meio, da crise dos recursos naturais energéticos e alimentares, das mudanças climáticas que atingem todo o planeta, se tornam mais impactantes em áreas periféricas, que além de acometidos pela pobreza, sequer tem os seus direitos civis atendidos. São nessas áreas periféricas que o poder público, em geral, está ausente e onde existe uma deficiência de estrutura capaz de oferecer à população oportunidades dignas de serviços, trabalho e lazer.

Nesta perspectiva, Kowarick (2000) expõe que não só a relação natureza e sociedade gera a vulnerabilidade, mas afirma que existe muita vulnerabilidade em relação a direitos básicos, na medida em que não só os sistemas públicos de proteção social foram sempre restritos e precários, como também, em anos recentes, houve desmonte de serviços e novas regulamentações que se traduziram em perda de direitos adquiridos.

Embasado nesse entendimento é que se reconhece que o estudo da vulnerabilidade envolve uma discussão ampla e relevante por ter um caráter multidisciplinar e indica que a suscetibilidade das pessoas a problemas e danos estão, principalmente, relacionadas ao conjunto das profundas transformações sociais, econômicas e ambientais que afetam, pelo mundo inteiro, as pessoas ou grupos de pessoas (KOWARICK, 2009).

Ademais, precisa compreender que a definição de vulnerabilidade socioambiental deve envolver uma percepção mais integrada das condições de vida de uma dada população. Deve-se olhar e perceber que os impactos advindos dos desastres naturais reconfiguram os cenários urbanos e impactam a forma e condição de vida das pessoas, mas também, que o contexto de construção da sociedade e do processo de expansão urbana nas cidades brasileiras, faz com que uma parte da população desconheça seus

direitos e experimentem de forma intensa uma sobreposição de desigualdades sociais: pobreza, segregação espacial, ausência de conforto urbano e, principalmente, os direitos à cidadania.

De acordo com o documento final da Conferência Mundial para a Redução de Desastres em Kobe (UN, 2005) é latente a necessidade de se desenvolver sistemas de indicadores de risco e vulnerabilidade nos níveis nacional e subnacional como forma de permitir aos tomadores de decisão um melhor diagnóstico das situações de risco e vulnerabilidade.

A utilização dos indicadores tem sido basilar na formulação de políticas públicas, pois possibilitam uma análise atual, conhecendo verdadeiramente a situação que se almeja modificar, como também o monitoramento temporal, permitindo um melhor acompanhamento das variáveis analisadas, uma melhor avaliação dos processos e planejamento de ações e, por conseguinte melhor utilização dos recursos investidos.

Deste modo, este artigo tem como objetivo sistematizar um conjunto de indicadores de vulnerabilidade socioambiental utilizando a análise de cluster baseada em similaridade textual.

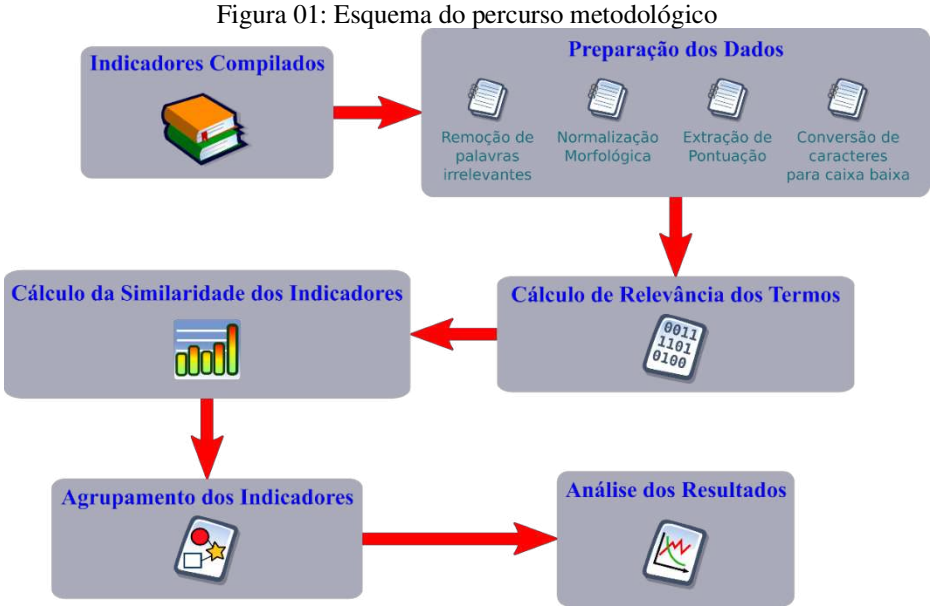
Para cada uma das etapas propostas, foram indicadas técnicas matemáticas capazes de executar e viabilizar tais recomendações. Ademais, foi utilizada a linguagem de programação *Python,* dado ao fato de, apesar de tratar-se de uma linguagem de programação de propósito geral, ela possui um ecossistema significativo para a mineração de dados, mineração textual e processamento de linguagem natural, o que facilita a implementação das técnicas sugeridas nas etapas da proposta metodológica.

É importante destacar que a utilização de técnicas matemáticas e da linguagem de programação, permite uma análise mais clara e livre de juízo de valor dos pesquisadores envolvidos.

2 Sistematização de um conjunto de indicadores: metodologia genérica

A metodologia proposta é constituída de seis etapas, descritas na Figura 01, as quais indicam quais os métodos e técnicas de devem ser utilizadas em cada uma delas. Ademais, a intenção é que se facilite o entendimento da proposição realizada, para que no tópico subsequente, possa-se compreender a aplicação realizada na sistematização de um conjunto de indicadores de vulnerabilidade socioambiental.

Figura 01: Esquema do percurso metodológico



Fonte: Elaboração própria (2020)

Com base na Figura do percurso metodológico proposto, segue a explicitação e o delineamento de cada uma das etapas que o compõe.

2.1) Indicadores Compilados

Esta etapa consiste na elaboração de um banco de dados composto por indicadores capazes de analisar o fenômeno a ser estudado. Para tanto, é necessário que se tenha conhecimento teórico acerca do referido fenômeno que se quer analisar, para que se tenha embasamento suficiente para se estabelecer critérios de escolha e de pesquisa dos indicadores que deverão ser compilados.

Após a consolidação do arcabouço teórico dos indicadores, sugere-se o uso de Mineração Textual (*Text data mining ou Knowledge discovery from textual databases*). Segundo Castro e Ferrari (2016) o referido termo foi firmado como alusão ao processo de extração de minerais valiosos a partir de uma mina. Assim, o processo de mineração textual envolve a "exploração de uma base de dados (mina) usando algoritmos (ferramentas) adequados para obtenção do conhecimento (minerais preciosos)" (CASTRO; FERRARI, 2016). Nesse contexto, o conjunto de dados a se analisar é um conjunto de textos, estruturados ou não.

Assim, o processo de mineração textual envolve a aplicação de técnicas que avaliam a sequência dos termos no contexto dos textos, fundamentalmente baseadas no Processamento de Linguagem Natural (PNL), que consiste da análise de dados textuais com base em conhecimentos relacionados com a forma de falar e escrever do ser humano. Dos possíveis conhecimentos importantes ao PNL, principalmente para o contexto desse trabalho, destacam-se:

1) Conhecimento Morfológico: refere-se ao conhecimento da estrutura, da forma e das inflexões das palavras;

2) Conhecimento Sintático: diz respeito ao conhecimento estrutural das listas de palavras e como elas podem ser combinadas para produzir sentenças;

3) Conhecimento Semântico: é o conhecimento do significado das palavras, independente do contexto. Ademais, também designa outros significados mais complexos, os quais podem ser obtidos pela combinação destas palavras.

Além destes três destacados, existem outros tipos de conhecimentos, como o pragmático, do discurso e do mundo, os quais tem importância em determinados escopos de aplicação. No entanto, para a proposta metodológica que aqui está sendo apresentada, estes foram desconsiderados, uma vez que os três em destaque e pontuados de forma mais específica, atendem ao objetivo.

Assim sendo, a etapas seguintes são ferramentas de mineração textual, aplicadas ao contexto metodológico que se apresenta.

2.2) Preparação dos Dados

Esta etapa é realizada visando preparar os dados para uma análise eficaz e consiste em: remoção de palavras irrelevantes, normalização morfológica, extração de pontuação e conversão de caracteres para caixa baixa.

3.2.1) Remoção de palavras irrelevantes

Esta etapa refere-se a limpeza de ruídos que podem atrapalhar a consistência da análise. Define-se como palavras irrelevantes (*stopwords*) as palavras que não traduzem o cerne do texto, tais como as preposições, pronomes, artigos, advérbios, e outras classes de palavras auxiliares.

3.2.2) Normalização Morfológica

No contexto de mineração de dados, normalização consiste no processo de transformação dos dados com o objetivo de facilitar sua análise. A sugestão nesta metodologia é a utilização da normalização morfológica, ou *stemming*, que, segundo Wives (2002), consiste no processo de eliminação de variações morfológicas de uma palavra ou frase, tais como prefixos e sufixos, reduzindo-a a seu radical ou termo próximo. A aplicação de normalização morfológica facilita a comparação entre termos, uma vez que possíveis variações oriundas de conjugações verbais, de gênero ou de número, podem dificultar a verificação de similaridade ou dissimilaridade.

3.2.3) Extração de pontuação e conversão de caracteres para caixa baixa

As palavras que possuem caracteres em caixa alta ou que possuem pontuação associada podem ser considerados como diferentes das palavras idênticas que não possuam estas características. Dessa forma, a extração de pontuação e a conversão de caracteres para caixa baixa, que também constituem

uma normalização dos termos avaliados, auxiliam na verificação de similaridade/dissimilaridade entre termos comparados.

2.3) Cálculo de relevância dos termos

Para esta etapa, sugere-se que os indicadores sejam representados por um modelo de espaço vetorial, no qual os textos analisados sejam convertidos em vetores numéricos por meio da análise da relevância dos termos que compõe os textos.

Cada documento ou texto que embase os indicadores será representado por um vetor de termos e cada termo possuirá um valor associado que indicará o grau de importância (definido como peso) desse documento.

É importante destacar que nem todas as palavras presentes em um documento possuem a mesma importância. Dessa forma, o grau de importância será proporcional à frequência do termo no texto (com exceção das palavras irrelevantes), e/ou entre textos que estejam sendo conjuntamente analisados. Ademais, também serão considerados importantes as palavras que compõem os títulos, por descreverem a ideia geral do documento, bem como os substantivos e complementos (SARKAR, 2016). Desta forma, o cálculo de relevância dos termos em relação ao texto em que está inserido pode basear-se: na frequência do termo, na análise estrutural do documento, ou na posição sintática de uma palavra.

Para esta proposta metodológica é utilizada a análise baseada na frequência do termo. Tal proposição é embasada na simplicidade de sua aplicação, principalmente por dispensar a utilização de técnicas mais avançadas de processamento de linguagem natural que as outras duas (análise estrutural do documento e posição sintática de uma palavra) exigem. Nessa proposta metodológica, é sugerida a utilização de uma combinação de duas medidas: a frequência dos termos (TF) e a frequência inversa dos documentos (IDF).

- A Freqüência do Termo, ou T*erm Frequency* (TF), representa a frequência absoluta de um termo em um texto ou documento. Nessa medida, basicamente é feita uma contagem das vezes em que um termo aparece no texto analisado, desconsiderando a contagem em outros textos e a quantidade total de termos do texto. A ideia é que termos que possuem uma pequena ocorrência dentro de um texto são menos relevantes que termos com maior ocorrência. No entanto, essa medida isolada pode levar a interpretações errôneas, como a de atribuir mesma relevância a uma palavra em um documento pequeno, quando comparada com outro termo em um documento grande, simplesmente porque ambos aparecem a mesma quantidade de vezes.
- A Freqüência Inversa de Documentos, ou I*nverse Document Frequency* (IDF), por sua vez, contabiliza a frequência de um termo com relação ao número de palavras que o texto analisado possui, além de considerar a incidência da palavra em outros documentos, o que pode destacar ainda mais a relevância de um termo.

Nessa proposta, a relevância de um termo (R_t) é, calculada como sendo o produto dessas duas medidas

$$R_t = TF \cdot IDF.$$

Daí, Sarkar (2016) citar esse método como sendo TF-IDF.

2.4) Cálculo da similaridade dos indicadores

Documentos podem ser representados como vetores, onde cada atributo representa a frequência na qual um determinado termo ocorre no documento. Então, para cada documento se terá um vetor com componentes e suas ponderações (TAN et al; 2009). Dessa forma, dois documentos serão similares se tiverem vetores com componentes proporcionalmente parecidos, mesmo que as componentes de um documento sejam maiores do que a do outro. Assim, o espaço vetorial considerado terá um eixo para cada termo.

Para determinar esta similaridade, sugere-se a utilização do método da similaridade dos cossenos, por ser uma das medidas mais comuns de semelhança de documentos (TAN et al, 2009). Nesse método, a similaridade é calculada pelo produto interno entre os dois vetores que representam os termos ou textos comparados, ou seja, é a medida do cosseno do ângulo entre os dois vetores. Assim, se a similaridade calculada for 1, significa que o ângulo entre os vetores é 0º ($cos\ 0^o = 1$) e, portanto, os documentos analisados são idênticos. Já se a similaridade calculada for 0, significa que o ângulo entre

os vetores é 90º (*cos 90º= 0*) e, portanto, os documentos analisados são completamente dissimilares. Assim sendo, a similaridade do cosseno entre dois documentos pode assumir qualquer valor entre 0, que indica completa dissimilaridade, e 1, que indica total similaridade.

Dessa forma, nessa proposta, a partir de uma comparação par a par entre todos os indicadores, se obterá uma matriz simétrica, *m* x *m*, chamada aqui de matriz de similaridade (M_s), que apresenta as medidas de similaridade entre todos os *m* indicadores analisados, da seguinte forma

$$\mathrm{M_s} = \begin{pmatrix} S_{11} & S_{12} & S_{13} & \cdots & S_{1n} \\ S_{21} & S_{22} & S_{23} & \cdots & S_{2m} \\ S_{31} & S_{32} & S_{33} & \cdots & S_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & S_{m3} & \cdots & S_{mm} \end{pmatrix},$$

em que os termos $s_{ij} = s_{ji}$ possuem valores entre 0 e 1, sendo 1 o valor de máxima similaridade, assumido somente pelos termos da diagonal principal (s_{ii}).

2.5) Agrupamento dos indicadores

Realizada a medida de similaridade, sugere-se utilizar de métodos para agrupar os objetos similares entre si. Para tanto, optou-se pelo método das *k*-médias *(k-means)*, devido a sua simplicidade e eficiência quando usado com um conjunto de dados numéricos, como a matriz de similaridade gerada na etapa anterior. O algoritmo *k*-médias toma como entrada o número de grupos *k*, e particiona o conjunto de *n* objetos em *k* grupos, de forma que a similaridade *intragrupo* seja alta e a similaridade *intergrupo* seja baixa (CASTRO, FERRARI, 2016).

Nesse algoritmo, primeiro é selecionado o número de grupos desejados. Cada dado é, então, atribuído a um centroide (ou baricentro) por proximidade, de forma que cada coleção de dados associados a um centroide forme um grupo. Formado um grupo, o centroide é atualizado, com base na posição dos dados que o compõe. Os passos de atribuição e repetição são, então, repetidos até que o algoritmo convirja, ou seja, que nenhum dado mude de grupo e/ou nenhum centroide seja alterado (TAN et al, 2009).

2.6) Análise dos resultados

Nesta etapa, se analisa a relevância dos indicadores inseridos em cada agrupamento obtido na etapa anterior. Para tanto, deve-se fazer uso do coeficiente de variação como parâmetro, uma vez que a medida do coeficiente de variação mostra o tamanho do espalhamento das medidas de similaridade de uma variável em relação a todas as outras do conjunto. Assim, quanto maior for o coeficiente de variação de uma variável, mais dissimilar esta variável será das demais que compõe o agrupamento.

Após a explicitação de cada uma das etapas da proposta metodológica, destaca-se que sua implementação pode ser utilizada a partir da utilização de *softwares* ou linguagens de programação, dentre as quais pode-se citar: SPSS®, *Sphinx*®, R, *Python*. A escolha deve ser pautada na facilidade do pesquisador em utilizar o *software*/linguagem que lhe mais for acessível.

Após a descrição das técnicas a serem utilizadas em cada uma das etapas, o tópico seguinte tratará de sua aplicação para a construção de um sistema de indicadores de vulnerabilidade socioambiental urbano.

3 Sistematização de um conjunto de indicadores de vulnerabilidade socioambiental: metodologia específica

Tomando como base a metodologia exposta anteriormente, neste tópico, será apresentado a aplicação para sistematização do conjunto de indicadores de vulnerabilidade socioambiental proposto por Vasconcelos (2019), utilizando a análise de cluster baseada em similaridade textual. Para tanto, optou-se por fazer uso da linguagem *Python*, por se tratar-se de uma linguagem de programação de propósito geral e possuir um ecossistema significativo para a mineração de dados, mineração textual e processamento de linguagem natural, o que facilita a implementação das técnicas sugeridas nas etapas da proposta metodológica.

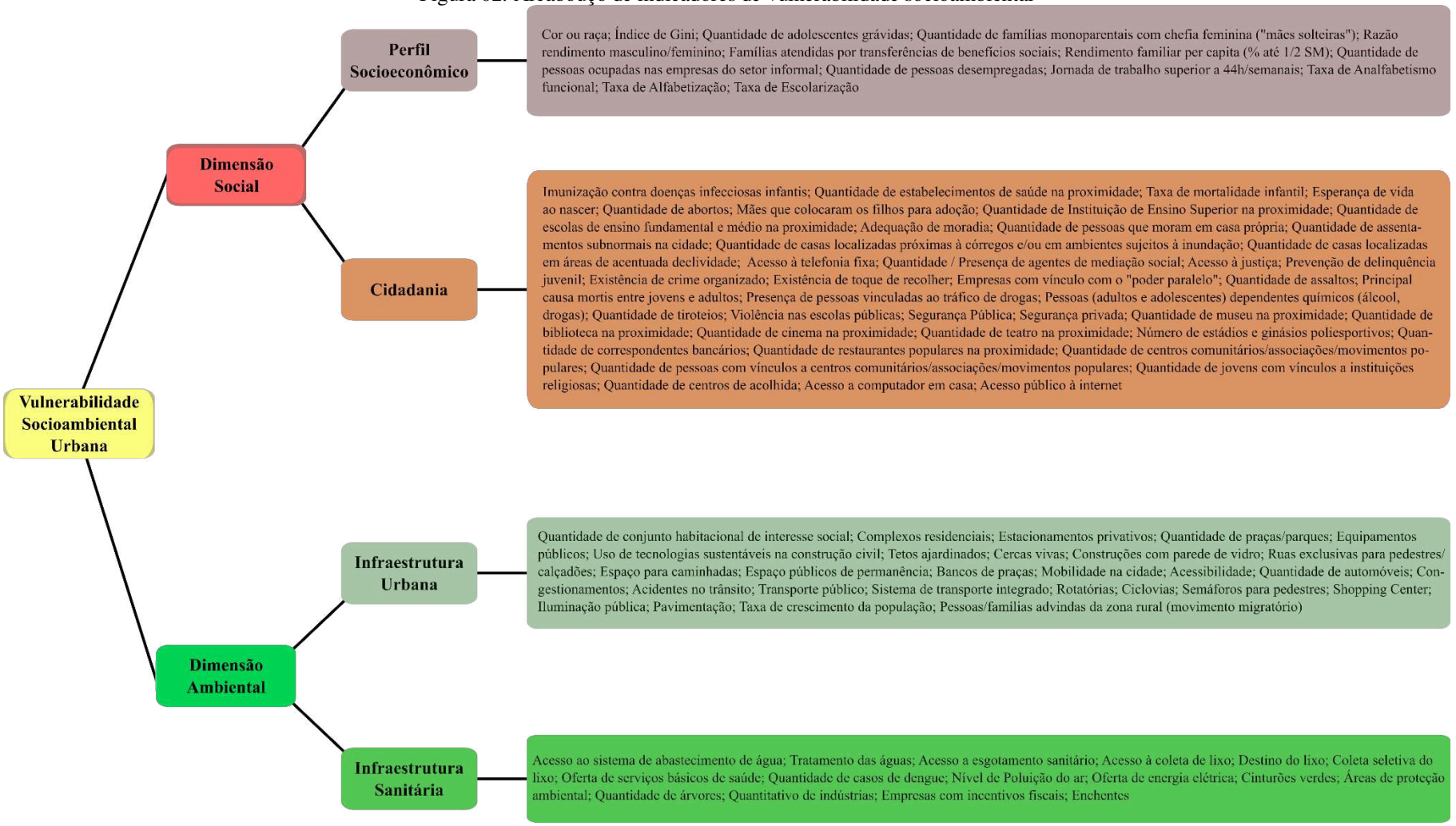
3.1) Compilação dos indicadores

Nesta fase, que é necessário se ter uma base de indicadores para realizar a mineração de dados, tomou-se como base, o arcabouço de 98 indicadores de vulnerabilidade socioambiental e apresentados na Figura 01.

Fazendo uma analogia com a definição de mineração de dados, o arcabouço escolhido é considerado a mina, as informações analisadas são as suas respectivas justificativas apresentadas por Vasconcelos (2019), as quais serão comparadas com base na análise de cluster baseado em similaridade textual. A escolha pela justificativa de cada indicador como dado textual a ser analisado, deve-se ao fato desta revelar a importância do indicador e sua relação com o fenômeno da vulnerabilidade socioambiental.

Dentre os conhecimentos sugeridos na metodologia indicada no tópico 2 deste artigo, optou-se aqui pela análise sintática que permitiu o conhecimento da estrutura das palavras que compõem cada justificativa dos indicadores. Também se fez uso da análise semântica das justificativas apresentadas por Vasconcelos (2019) dos indicadores de vulnerabilidade socioambiental, a qual permitiu o conhecimento do significado das palavras de cada justificativa.

Figura 02: Arcabouço de indicadores de vulnerabilidade socioambiental



Fonte: Adaptado de Vasconcelos (2019)

3.2) Preparação dos dados

Nesta etapa foi realizada a preparação dos dados para a análise das justificativas.

3.2.1) Remoção de palavras irrelevantes

Nesta etapa foram retiradas as palavras irrelevantes. Os textos utilizados como base foram as justificativas dos indicadores de vulnerabilidade socioambiental apresentadas por Vasconcelos (2019). Aqui, foram retiradas as preposições, pronomes, artigos, advérbios e outras classes de palavras auxiliares. As palavras irrelevantes formam uma lista de palavras conhecida como *stoplist*. As palavras que compõem a *stoplist* dificilmente são utilizadas em uma consulta.

Nesse trabalho usou-se o corpus de palavras irrelevantes para o português disponível na base de dados da biblioteca NLTK (2017) da linguagem *Python*.

3.2.2) Normalização Morfológica

Para realizar essa etapa foi necessário um algoritmo específico de normalização morfológica, que corresponde a um banco de palavras (dicionário morfológico) que permite a comparação entre as palavras. São poucos os existentes na língua portuguesa, menos ainda para o português brasileiro. Por este motivo, neste estudo, fez-se uso do algoritmo RSLP desenvolvido na UFRGS por Orengo e Huyck (2001), que foi desenvolvido para normalização morfológica no português brasileiro. A implementação usada nesse trabalho está disponível na biblioteca NLTK (2017) da linguagem *Python.*

3.2.3) Extração de pontuação e conversão de caracteres para caixa baixa

As palavras que possuem caracteres em caixa alta ou que possuem pontuação associada podem ser considerados como diferentes das palavras idênticas que não possuam estas características. Assim sendo, nesta fase, visando obter uma mineração dos textos de forma mais clara e precisa, foram excluídos os sinais de pontuação existentes nas justificativas dos indicadores, bem como todos os caracteres em caixa alta foram convertidos para caixa baixa.

3.3) Cálculo de relevância dos termos

Para este estudo, os indicadores serão representados por um modelo de espaço vetorial, no qual as justificativas de cada indicador, que serão os textos analisados, são transformados em vetores por meio da análise da relevância dos termos que compõe os textos.

O processo de vetorização e cálculo de relevância dos termos usados nesse trabalho, TF-IDF, é realizado usando a implementação disponível na biblioteca *Scikit-Learn*, da linguagem *Python*.

3.4) Cálculo da similaridade dos indicadores

Para o cálculo da similaridade dos indicadores foi realizado uma comparação par a par de todos os indicadores que compõem cada tema. Como pode ser observado na Figura 01, o arcabouço teórico dos indicadores é composto por duas dimensões (social e ambiental), as quais estão subdividas em temas e constituídas pelos indicadores. Assim sendo, a análise de similaridade foi realizada com base no texto de justificativa de cada indicador.

Analisando as matrizes de similaridade (Anexo 02), e utilizando o método de similaridade do cosseno, verifica-se que quanto mais próximo de 1, maior será a similaridade entre os indicadores. Do contrário, quanto mais próximo de 0, maior a dissimilaridade entre os indicadores.

Assim, na comparação par a par entre todos os indicadores que compõem um tema, tem-se que quanto mais próximo de 01 resultou a comparação entre eles, maior o grau de similaridade e quanto mais próximo de zero, mais dissimilares os indicadores de vulnerabilidade socioambiental serão.

O grau de similaridade encontrado entre os indicadores e apresentados nas matrizes, viabilizará o agrupamento dos indicadores.

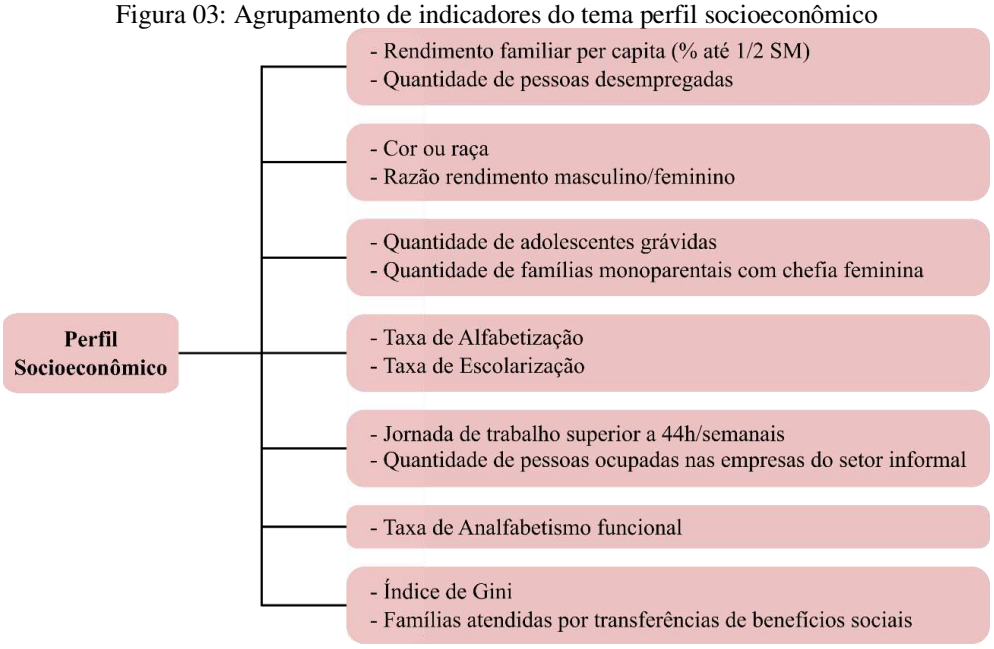
3.5) Agrupamento dos indicadores

Após a obtenção da análise de cluster baseada em similaridade textual tem-se os agrupamentos dos indicadores. Para cada tema foi elaborada uma representação gráfica, demonstrando quais indicadores foram considerados mais similares a partir da comparação das suas justificativas. Os indicadores que, a partir de suas justificativas, apresentaram similaridades estarão em um mesmo

agrupamento. Os indicadores que não apresentaram similaridades com outro indicador do seu tema, estarão apresentados isoladamente.

A Figura 03 apresenta o agrupamento de indicadores que compõem o tema **Perfil Socioeconômico**. Este tema é composto por 14 indicadores.

Figura 03: Agrupamento de indicadores do tema perfil socioeconômico

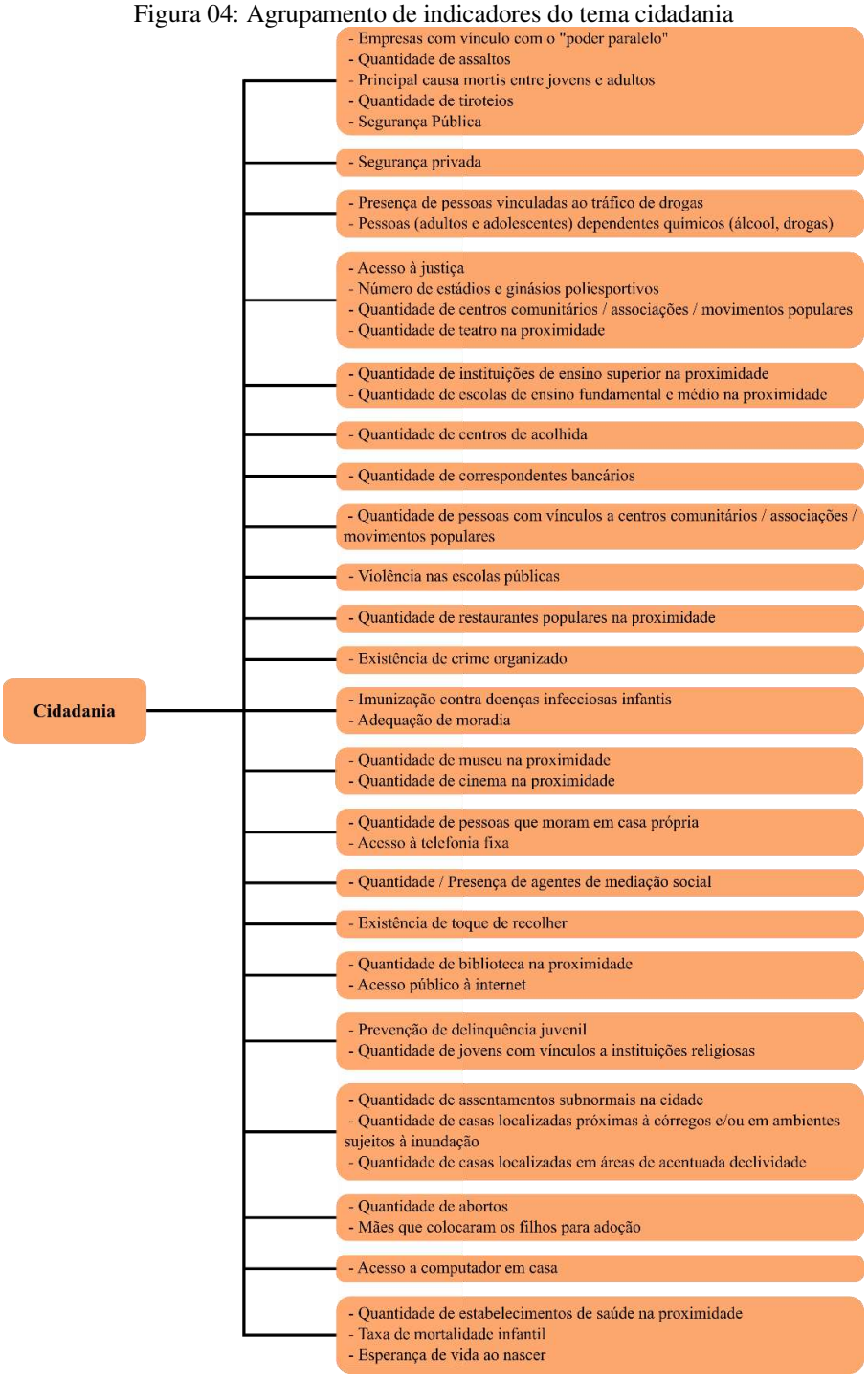


Fonte: Elaboração própria (2020)

Após a análise da matriz de similaridade, verificou-se que dos 14 indicadores que compõem este tema, resultaram 07 agrupamentos: 06 agrupamentos compostos por dois indicadores que se apresentaram similares e 01 indicador isolado, o qual se apresentou dissimilar de todos os outros indicadores.

A Figura 04 apresenta o agrupamento de indicadores que compõem o tema **Cidadania**. Este tema é composto por 41 indicadores.

Figura 04: Agrupamento de indicadores do tema cidadania

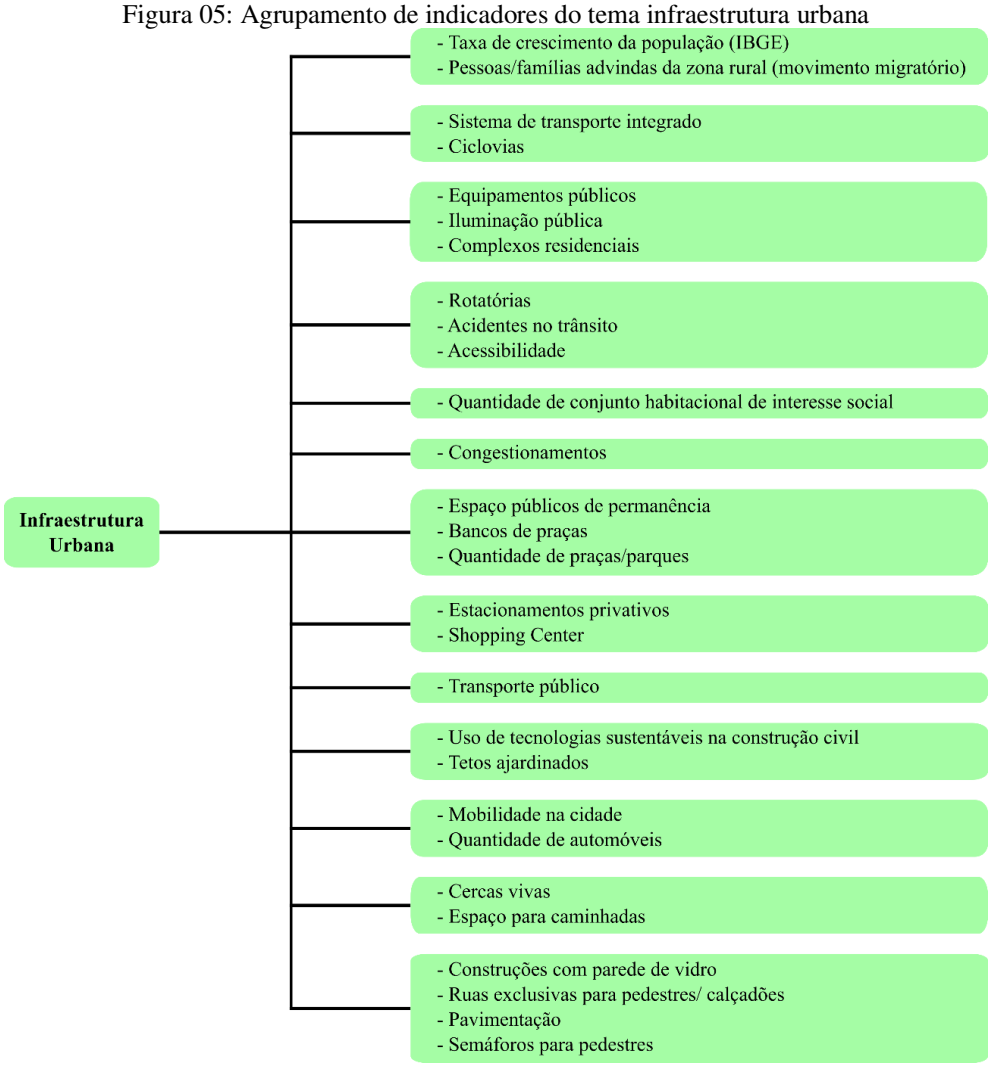


Fonte: Elaboração própria (2020)

Após a análise da matriz de similaridade, verificou-se que dos 41 indicadores que compõem este tema, resultaram 22 agrupamentos: 01 agrupamento composto por cinco indicadores; 01 agrupamento composto por 04 indicadores, 02 agrupamentos compostos por três indicadores; 08 agrupamentos que apresentam dois indicadores e 10 indicadores que são dissimilares, a partir da comparação de suas justificativas.

A figura 05 apresenta o agrupamento de indicadores que compõem o tema **Infraestrutura Urbana**, o qual é constituído por 32 indicadores.

Figura 05: Agrupamento de indicadores do tema infraestrutura urbana

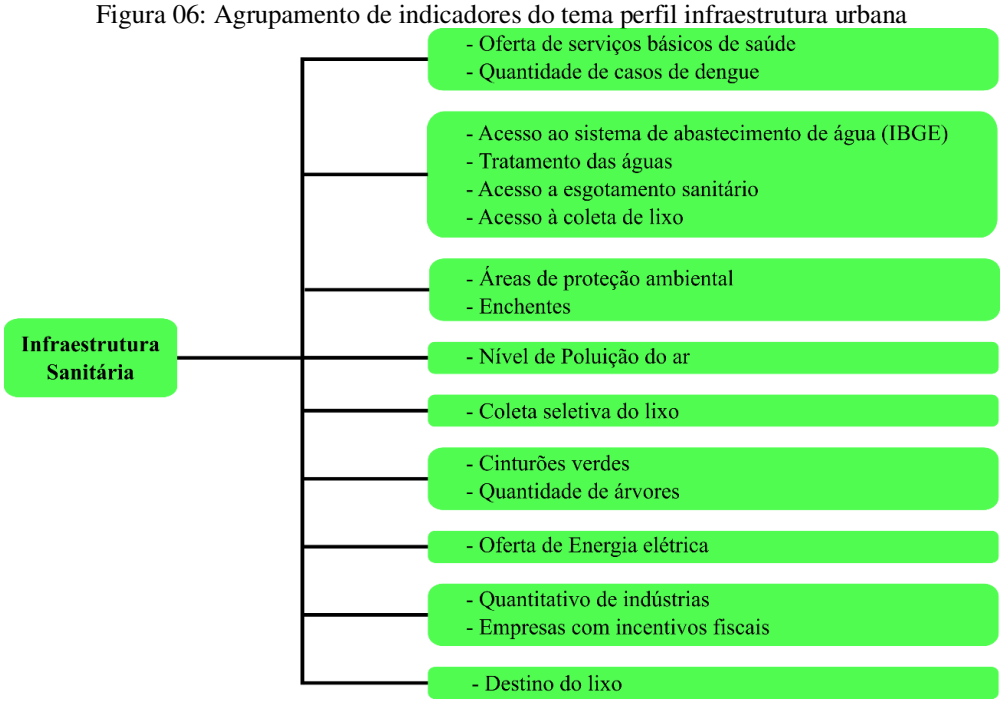


Fonte: Elaboração própria (2020)

Dos 32 indicadores resultou-se em 15 agrupamentos, dos quais 02 agrupamentos são constituídos por quatro indicadores; 04 agrupamentos compostos por três indicadores; 03 agrupamentos compostos por três indicadores; 06 indicadores são dissimilares a partir de suas justificativas e não se apresentam agrupados.

A figura 06 explicita os agrupamentos que constituem o perfil **Infraestrutura sanitária.** Composto por 17 indicadores, a matriz de similaridade permitiu a obtenção de 09 agrupamentos.

Figura 06: Agrupamento de indicadores do tema perfil infraestrutura urbana



Fonte: Elaboração própria (2020)

Conforme se pode observar, o tema em questão é comporto por 09 agrupamentos: 03 agrupamentos constituído por três indicadores, 02 agrupamentos constituídos por dois indicadores e 04 indicadores que não se agruparam com outros indicadores por suas justificativas não apresentarem similaridade com nenhum outro indicador.

Com base nas 04 figuras apresentadas neste tópico, pode-se observar uma redução significativa de variáveis quando os indicadores passaram a ser agrupados a partir da similaridade de suas justificativas. Dos 98 indicadores obteve-se 46 agrupamentos, os quais compõem a sistematização de um conjunto de indicadores de vulnerabilidade socioambiental. A redução no quantitativo de variáveis e a identificação da similaridade entre os indicadores, proporcionam maior flexibilidade e abrangência da análise da vulnerabilidade socioambiental, como pode ser constatado no tópico seguinte.

3.6) Análise dos resultados

Após a aplicação da metodologia proposta no item 2 e aplicadas ao conjunto de indicadores de vulnerabilidade socioambiental proposto por Vasconcelos (2019), tem-se, no quadro 02 abaixo, a sistematização de um conjunto de indicadores de vulnerabilidade socioambiental utilizando a análise de cluster baseada em similaridade textual, composto por 46 variáveis e suas similaridades, o qual pode ser observado no Quadro 02 abaixo.

Quadro 02: Sistema de Vulnerabilidade Socioambiental Urbano

VULNERABILIDADE SOCIOAMBIENTAL		
DIMENSÃO	**INDICADORES**	**AGRUPAMENTOS**
SOCIAL	**Perfil Socioeconômico**	1. Rendimento familiar per capita (% até ½ SM); Quantidade de pessoas desempregadas
		2. Cor ou raça; Razão rendimento masculino/feminino
		3. Quantidade de adolescentes grávidas; Quantidade de famílias monoparentais com chefia feminina
		4. Taxa de Alfabetização; Taxa de Escolarização
		5. Jornada de trabalho superior a 44h/semanais; Quantidade de pessoas ocupadas nas empresas do setor informal
		6. Taxa de analfabetismo
		7. Índice de Gini; Famílias atendidas por transf. de benefícios sociais
	Cidadania	8. Acesso à Justiça; Nº de estádios e ginásios poliesportivos; Quantid. De centros comunitários/associações/movimentos populares; Quantidade de teatro na proximidade
		9. Quant. de IES na proximidade; Quant. de escolas de ensino fundamental e médio na proximidade
		10. Quantidade de centros de acolhida
		11. Quantidade de correspondentes bancários
		12. Quantidade de pessoas com vínculos a centros comunitários/ associações/movimentos populares
		13. Existência de crime organizado
		14. Imunização contra doenças infecciosas infantis; Adequação de moradia
		15. Quantidade de museu na proximidade; Quantidade de cinema na proximidade
		16. Quantidade de pessoas que moram em casa própria; Acesso à telefonia fixa
		17. Quantidade/presença de agentes de mediação social
		18. Existência de toque de recolher
		19. Quantidade de biblioteca na proximidade; Acesso público à internet
		20. Prevenção de delinquência juvenil; Quant. de jovens com vínculos a instituições religiosas
		21. Quant. de assentamentos subnormais na cidade; Quant. de casas localizadas próximas à córregos e/ou em ambientes sujeitos à inundação; Quant. de casas localizadas em áreas de acentuada declividade
		22. Quantidade de abortos; Mães que colocaram os filhos para adoção
		23. Acesso a computador em casa
		24. Quant. de estabelecimentos de saúde na proximidade; Taxa de mortalidade infantil; Esperança de vida ao nascer
AMBIENTAL	**Infraestrutura Urbana**	25. Taxa de crescimento da população; Pessoas/famílias advindas da zona rural (movimento migratório)
		26. Sistema de transporte integrado; Ciclovias
		27. Equipamentos públicos; Iluminação pública; Complexos residenciais
		28. Rotatórias; acidentes no trânsito; Acessibilidade
		29. Quant. de conjuntos habitacional de interesse social
		30. Congestionamentos
		31. Espaços públicos de permanência; Bancos de praça; Quant. de praças/parques
		32. Estacionamentos privativos; Shopping center
		33. Transporte público
		34. Uso de tecnologias sustentáveis na construção civil; Tetos ajardinados
		35. Mobilidade na cidade; Quant. de automóveis
		36. Cercas vivas; Espaço para caminhadas
		37. Construções com parede de vidro; Ruas exclusivas para pedestres/calçadões; Pavimentação; Semáforos para pedestres
	Infraestrutura Sanitária	38. Oferta de serviços básicos de saúde; Quant. de casos de dengue
		39. Acesso ao sistema de abastecimento de água; Tratamento de águas; Acesso a esgotamento sanitário; Acesso à coleta de lixo
		40. Áreas de proteção ambiental; Enchentes
		41. Nível de poluição do ar
		42. Coleta seletiva do lixo
		43. Cinturões verdes; Quantidade de árvores
		44. Oferta de energia elétrica
		45. Quant. de indústrias; Empresas com incentivos fiscais
		46. Destino do lixo

Fonte: Elaboração própria (2020)

Com base no quadro 02 construído a partir do resultado do agrupamento dos indicadores, pode-se observar que cada agrupamento foi realizado em função da similaridade das justificativas de cada indicador. Assim, entende-se que um indicador, por similaridade, é capaz de oferecer entendimento e subsidiar análises dos demais indicadores que compõem o agrupamento. Dessa forma, constata-se que

a análise de todos os indicadores de um agrupamento viabilizará a análise das inter-relações existentes entre os indicadores, proporcionando uma análise mais ampla e complexa.

Com a finalização da aplicação da proposta metodológica para a sistematização de um conjunto de indicadores, pode-se comprovar que sua aplicação é exequível e viável dada a sua praticidade e objetividade para análise do fenômeno em questão.

Quanto a aplicação da metodologia proposta, voltada para o fenômeno da vulnerabilidade socioambiental, tem-se como resultado, um sistema capaz de analisar tal fenômeno, de forma abrangente, integrada e permitindo flexibilidade ao pesquisador, ao passo que permite tanto a análise dos agrupamentos integralizados ou dos indicadores que o represente dentro do agrupamento.

Assim sendo, tanto a proposta metodológica aqui apresentada, quanto a sua aplicação no contexto da vulnerabilidade socioambiental, constitui avanços importantes e significativos nos estudos que envolve indicadores, sistemas de indicadores e vulnerabilidade socioambiental.

4 CONSIDERAÇÕES

Os desafios enfrentados com a elaboração de indicadores e sistemas de indicadores são muitos, pois trata-se de revelar uma realidade complexa, de uma forma clara e objetiva. Muitas são as iniciativas na construção de um sistema de indicadores de vulnerabilidade socioambiental, mas, em sua maioria, os indicadores não se apresentam integrados. Esta lacuna foi suprida com este estudo, ao passo que se utilizou medidas de similaridade e técnicas de agrupamento de dados para integrar indicadores e permitir uma análise mais abrangente.

O alcance do objetivo deste estudo, qual seja, sistematizar um conjunto de indicadores de vulnerabilidade socioambiental utilizando a análise de cluster baseada em similaridade textual, enfatizou que além de sua contribuição teórica e seu caráter inovador, constitui um avanço significativo nos estudos acerca da temática.

A sua exequibilidade voltada para o fenômeno da vulnerabilidade socioambiental, permite compreender a sua eficiência na redução do número de indicadores, sem ser reducionista. Isto é possível, em função dos agrupamentos encontrados, os quais exprimem o contexto e a interligação existente entre um indicador e os demais que compõem o agrupamento que pertence.

Deste modo, mesmo se escolhendo ou só sendo possível analisar um indicador de um determinado agrupamento, este revelará informações secundárias dos demais indicadores, resultado da identificação do grau de similaridade com os demais.

Assim sendo, este estudo atende ao seu objetivo inicial e não pretende com isto esgotar as possibilidades de sistematização de um conjunto de indicadores, mas direcionar novos estudos em que novas técnicas possam ser aplicadas. Também não é de interesse deste estudo, limitar aqui a construção de um sistema de indicador para análise da vulnerabilidade socioambiental, mas sugerir que novos indicadores e novos temas possam ser agregados aos que aqui foram sistematizados, e que, assim, se consiga se obter um sistema que atenda as multifaces que envolve o referido fenômeno.

Neste contexto, pretende-se aqui, oferecer subsídios para que a metodologia sugerida possa ser aplicada para analisar outros fenômenos que não apenas a vulnerabilidade socioambiental, mas que possa abranger outros fenômenos.

Ademais, sugere-se também que o resultado da aplicação que gerou a sistematização de um conjunto de indicadores subdivididos em temas e dimensões para o fenômeno da vulnerabilidade socioambiental, possa ser aplicado em uma cidade de médio ou grande porte, avaliando a sua praticidade e exequibilidade.

5 REFERÊNCIAS

ACSERALD, Henri. **Vulnerabilidade ambiental, processos e relações**. Comunicação ao II Encontro Nacional de Produtores e Usuários de Informações Sociais, Econômicas e Territoriais, FIBGE, Rio de Janeiro, 24/08/2006.

ALMEIDA, Lutiane Queiroz de. **Vulnerabilidade Socioambiental de rios urbanos.** Bacia hidrográfica do rio Maranguapinho. Região metropolitana de Fortaleza, Ceará. 2010. 278 f. Tese de Doutorado em Geografia – Instituto de Geociência e Ciência Exatas, Unesp, Rio Claro, 2010.

ALVES, Humberto Prates da Fonseca **Vulnerabilidade socioambiental na metrópole paulistana**: uma análise sociodemográfica das situações de sobreposição espacial de problemas e riscos sociais e ambientais. Revista Brasileira de Estudos de População. vol.23; nº 01; São Paulo Janeiro/Junho 2006.
ALVES, H. P. et al. **Dinâmicas de urbanização na hiperperiferia da metrópole de São Paulo:** análise dos processos de expansão urbana e das situações de vulnerabilidade socioambiental em escala intraurbana. Revista Brasileira de estudos populacionais. Rio de Janeiro, v. 7, n. 1, p. 141-159, jan./jun. 2010.
BLAIKIE, P. M.; CANNON, T.; DAVIS, I.; WISNER, B. **At risk:** natural hazards, people´s vulnerability, and disasters. London: Routledge, 1994.
BRAGA, Tania Moreira; OLIVEIRA, Elzira Lucia De; GIVISIEZ, Gustavo Henrique Naves. Avaliação de metodologias de mensuração de risco e vulnerabilidade social a desastres naturais associados à mudança climática. **XV encontro nacional de estudos populacionais,** Caxambú- MG, 2006. Disponível em: <www.abep.org.br/publicacoes/index.php/anais/article/download/1615/1578>. Acesso em: 23 maio 2017.
CASTRO, L. N; FERRARI, D. G. **Introdução à Mineração de Dados**: conceitos básicos, algoritmos e aplicações. Editora Saraiva. 1ª edição. São Paulo, 2016.
CUTTER, S. L. **Vulnerability to environmental hazards.** Progress in human geography 20, 4, 529-539, 2003.
CUTTER, S. L. A ciência da vulnerabilidade: modelos, métodos e indicadores. **Revista Crítica de Ciências Sociais**, v. 93, n. 1, p. 59-70, jun. 2011.
D´ERCOLE, R. **Social vulnerability to environmental hazards.** Social Science Quarterly 84(1): 242-261, 2009.
GAMBA, C; RIBEIRO, W. C., **Indicador e avaliação da vulnerabilidade socioambiental no município de São Paulo**, GEOUSP - Espaço e Tempo, São Paulo, Nº 31 Especial, 2012
GUIMARÃES JRS, JANNUZZI PM. IDH, indicadores sintéticos e suas aplicações em políticas públicas. **Revista Brasileira de Estudos Urbanos e Regionais (ANPUR)** 2005;7(1):73-90.
HOGAN, D.J.; MARANDOLA JUNIOR, E. Towards an interdisciplinary conceptualization of vulnerability. **Population, Space and Place**, n. 11, p. 455-471, nov. 2005.
IBGE. **Censo demográfico brasileiro 2010.** Disponível em: <http://www.ibge.gov.br>. Acesso em: 22 de janeiro de 2017.
JACOBI, Pedro. **Cidade e meio ambiente: percepções e práticas em São Paulo**. Editora Annablume, 2ª edição. São Paulo, 2006.
KOWARICK, Lúcio. **Escritos Urbanos**. Editora 34. 1ª edição. São Paulo. 2000.
KOWARICK, Lúcio. **Viver em risco: Sobre a vulnerabilidade socioeconômica e civil.** Editora 34. São Paulo. 2009.
KUHLICKE, Christian; SCOLOBIG, Anna; TAPSELL, Sue; STEINFÜHRER, Annett; DE MARCHI, Bruna, "Contextualizing Social Vulnerability: Findings from case studies across Europe", ***Natural Hazards***, 58(2), 789-810, 2011.
MAIOR, M. M. S. **Vulnerabilidade socioambiental e expansão urbana**: uma proposta metodológica para análise da cidade de João Pessoa-PB. 2014. Tese (Doutorado em Recursos Naturais). Universidade Federal de Campina Grande, Campina Grande/PB.
NLTK 3.2.4 DOCUMENTATION. **Natural language toolkit**. Disponível em: <http://www.nltk.org/>. Acesso em: 19 abr. 2017.
ORENGO, Viviane Moreira; HUYCK, Christian. A Stemming Algorithm for the Portuguese Language. HTTP://WWW.INF.UFRGS.BR. **Rslp stemmer (removedor de sufixos da lingua portuguesa)**. Disponível em: <http://www.inf.ufrgs.br/~viviane/rslp/>. Acesso em: 25 abr. 2017.
SARKAR, Dipanjan. **Text Analytics with Python**. Editora Apress. 1ª Edição. Bangalore – Índia. 2016.
TAN, P.; STEINBACH, M.; KUMAR, V.; **Introdução ao DATAMINING Mineração de Dados**. Editora Ciência Moderna Ltda. Rio de Janeiro, 2009.
UNDP (2004). **Reducing desaster risk: a challenge for development, a global report.** UNDP Bureau for Crisis Prevention and Recovery. New York: UNDP. 2004.
VAN BELLEN, H.M. **Indicadores de sustentabilidade**: uma análise comparativa. Editora FGV. Rio de Janeiro. 2005.

VASCONCELOS, A. C. F.; CÂNDIDO, G. A.; FREIRE, E.M.X. Vulnerabilidade Socioambiental: proposição de temas e indicadores para cidades brasileiras. **Revista Gaia Scientia**, 2019; Vol. 13(2); p.1-18.
VASCONCELOS, A. C. F. INDICADORES DE VULNERABILIDADE SOCIOAMBIENTAL: proposição de framework e aplicação na cidade de Natal - RN. 2019. **Tese** (Doutorado em Desenvolvimento e Meio Ambiente). Universidade Federal do Rio Grande do Norte, Natal/RN.
WIVES, L. **Tecnologias de descoberta de conhecimento em textos aplicadas à inteligência competitiva**. EQ-069, PPGC-UFRGS, 2002.