



Dissertação de Mestrado

**Representação, visualização e análise de grandes
volumes de dados urbanos espaço-temporais de
segurança pública**

Tiago Paulino Santos

Orientador:

Thales Miranda de Almeida Vieira

Maceió, Dezembro de 2023

Tiago Paulino Santos

Representação, visualização e análise de grandes volumes de dados urbanos espaço-temporais de segurança pública

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Thales Miranda de Almeida Vieira

Maceió, Dezembro de 2023

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

S237r Santos, Tiago Paulino.

Representação, visualização e análise de grandes volumes de dados urbanos espaço-temporais de segurança pública / Tiago Paulino Santos. – 2023.

42 f. : il.

Orientador: Thales Miranda de Almeida Vieira.

Dissertação (mestrado em informática) - Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2023.

Bibliografia: f. 41-42.

1. Segurança pública. 2. Ciência de dados. 3. Análise de dados espaço-temporais. 4. Dados Urbanos. 5. *Big Data*. I. Título.

CDU: 004:351.75



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
Av. Lourival Melo Mota, S/N, Tabuleiro do Martins, Maceió - AL, 57.072-970
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO (PROPEP)
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Folha de Aprovação

TIAGO PAULINO SANTOS

**REPRESENTAÇÃO, VISUALIZAÇÃO E ANÁLISE DE GRANDES VOLUMES DE
DADOS URBANOS ESPAÇO-TEMPORAIS DE SEGURANÇA PÚBLICA**

**REPRESENTATION, VISUALIZATION AND ANALYSIS OF LARGE VOLUMES OF
SPATIO-TEMPORAL URBAN PUBLIC SAFETY DATA**

Dissertação submetida ao corpo docente do
Programa de Pós-Graduação em Informática
da Universidade Federal de Alagoas e
aprovada em 25 de janeiro de 2024.

Banca Examinadora:



Documento assinado digitalmente
THALES MIRANDA DE ALMEIDA VIEIRA
Data: 25/01/2024 15:19:01-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. THALES MIRANDA DE ALMEIDA VIEIRA
UFAL – Instituto de Computação
Orientador



Documento assinado digitalmente
RIAN GABRIEL SANTOS PINHEIRO
Data: 25/01/2024 15:35:43-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. RIAN GABRIEL SANTOS PINHEIRO
UFAL – Instituto de Computação
Examinador Interno



Documento assinado digitalmente
LUIS GUSTAVO NONATO
Data: 25/01/2024 15:31:33-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. LUIS GUSTAVO NONATO
USP – Universidade de São Paulo
Examinador Externo

Dedico este trabalho ao meu filho Arthur Gabriel Almeida Paulino, minha esposa, Karla Regina, minha avó, Maria dos Anjos e minha mãe Roseanna Paulino.

Agradecimentos

Gostaria de expressar minha profunda gratidão a todas as pessoas que contribuíram para a realização deste trabalho acadêmico. Cada passo, cada desafio superado, foi possível graças ao apoio e inspiração que recebi ao longo dessa jornada.

Primeiramente, dedico este trabalho à minha mãe, Roseanna, cujo amor incondicional e apoio constante foram a força motriz por trás de cada conquista. Sua dedicação à minha educação e bem-estar moldou o caminho que trilhei até aqui.

A minha amada esposa, Karla, merece uma gratidão especial. Seu apoio incondicional e compreensão durante os momentos desafiadores foram a âncora que me manteve firme diante das adversidades. Esta obra é dedicada a você, como uma expressão do profundo amor e gratidão que sinto.

Quero, neste momento, destacar e valorizar todos os pesquisadores que, como eu, enfrentam o desafio adicional de equilibrar o cuidado com filhos autistas. Um reconhecimento especial vai para o meu filho Arthur, um ser iluminado, autista, cuja presença em minha vida trouxe aprendizado e inspiração únicos. Esta obra é dedicada a ele.

Ao Professor Thales Vieira, sou grato pela orientação sólida e conhecimento valioso compartilhado durante este processo acadêmico. Sua paciência e incentivo foram fundamentais para o meu desenvolvimento como estudante e pesquisador.

Não posso deixar de mencionar o Professor Rômulo Nunes da UFAL-Arapiraca, cujo comprometimento com a excelência acadêmica e entusiasmo pela pesquisa científica foram fontes constantes de inspiração.

Finalmente, gostaria de prestar uma dedicatória póstuma ao saudoso amigo José Nicodemus. Sua influência e crença no meu potencial abriram portas para novos horizontes no mundo da ciência de dados. Agradeço por sua orientação e apoio contínuo, mesmo além da sua presença física.

Que este trabalho seja uma pequena contribuição ao campo da ciência, inspirando futuras gerações de pesquisadores, assim como eu fui inspirado por aqueles que me cercam.

Muito obrigado a todos.

*"O conhecimento é a única riqueza que cresce quando compartilhada."
(Sócrates)*

Resumo

As transformações tecnológicas advindas da transformação digital de órgãos públicos e da internet das coisas têm possibilitado a geração e coleta de grandes volumes de dados, que podem ser posteriormente utilizados em análises precisas. Para a segurança pública, a aplicação de novas tecnologias nas atividades de análise criminal, patrulhamento e repressão ao crime trazem inúmeras possibilidades também, como conseguir visualizar de maneira profunda o desempenho das ações da corporação, monitorar a atividade criminosa, compreender padrões e buscar alternativas para implementar melhores políticas de segurança. *Big Data* também já não é uma realidade distante das corporações de segurança pública. Dispositivos de rastreamento, câmeras de vigilância, sistemas de monitoramento, sistemas de atendimento e muitas outras fontes de informação já provêm um grande volume de dados que precisam ser devidamente tratados para que seja possível obtenção de conhecimento relevante. Nos últimos anos, muitos trabalhos científicos têm proposto o uso de algoritmos de Aprendizado de Máquina para reconhecer padrões espaciais e temporais de crimes. Nesse contexto, propomos neste trabalho uma ferramenta de análise visual de dados urbanos espaço-temporais, e um novo algoritmo para detecção de manchas criminais. Estas soluções foram validadas usando bases de dados da Polícia Militar do Estado de Alagoas (PMAL) em estudos de caso onde o objetivo era analisar dados espaço-temporais de crimes e de patrulhamento. Os resultados desta pesquisa terão importância não só do ponto de vista científico, mas também poderão ser aproveitados pela PMAL para otimizar seus processos de tomada de decisão. Um diferencial deste trabalho é a abordagem de fontes de dados de crimes e de patrulhamento na mesma ferramenta de visualização.

Palavras-chaves: Segurança Pública, Ciência de Dados, Análise de dados espaço-temporais, Dados Urbanos, *Big Data*.

Abstract

Technological transformations arising from the digital transformation of public bodies and the internet of things have enabled the generation and collection of large volumes of data, which can later be used in precise analyses. For public security, the application of new technologies in criminal analysis, patrolling and crime repression activities also bring countless possibilities, such as being able to visualize in depth the performance of the corporation's actions, monitor criminal activity, understand patterns and seek alternatives to implement better security policies. *Big Data* is also no longer a distant reality for public security corporations. Tracking devices, surveillance cameras, monitoring systems, customer service systems and many other sources of information already provide a large volume of data that needs to be properly processed in order to obtain relevant knowledge. In recent years, many scientific works have proposed the use of Machine Learning algorithms to recognize spatial and temporal patterns of crimes. In this context, we propose in this work a visual analysis tool for spatio-temporal urban data, and a new algorithm for detecting crime spots. These solutions were validated using databases from the Military Police of the State of Alagoas (PMAL) in case studies where the objective was to analyze spatio-temporal data on crimes and patrolling. The results of this research will be important not only from a scientific point of view, but can also be used by PMAL to optimize its decision-making processes. A difference of this work is the approach of crime and patrol data sources in the same visualization tool.

Keywords: Public Security; Data Science; Spatiotemporal data analysis; Urban data; Big Data.

Lista de Figuras

2.1	Pseudocódigo da <i>KD-Tree</i>	7
2.2	Exemplo de estrutura de organização dos nós de uma <i>Kd-Tree</i>	7
2.3	Pseudocódigo de construção da <i>Balltree</i>	8
2.4	Primeiro particionamento de pontos em um plano definidos numa <i>Ball-Tree</i> . . .	9
2.5	Pseudocódigo da busca de nós vizinhos na estrutura de <i>Balltree</i>	10
2.7	Subdivisão de uma rua (segmento de reta), de maneira a localizar a esquina mais próxima (ponto vermelho), a partir do ponto mais próximo da ocorrência no espaço vetorial V indicado pelos pontos. O ponto amarelo indica o ponto no espaço vetorial V mais próximo do ponto da ocorrência.	11
2.8	Representação da matriz de adjacência de um grafo.	12
2.9	Representação do Pseudocódigo de Djiskstra.	13
2.10	Pseudocódigo de ordenação de um vetor de Fiedler.	14
2.11	Grafo de Esquinas de Maceió/AL ordenado segundo o vetor de Fiedler.	15
2.12	Metodologia de visualização do Cripav para análise espacial e temporal de crimes, relacionando esses padrões à infraestrutura e sua mudança no tempo (Garcia-ZANABRIA et al., 2021).	16
4.1	Parte de um mapa da cidade de Maceió. Pontos menores em escala de cor ascendendo do azul para o vermelho indicam maior densidade de crimes. Círculos vermelhos indicam os <i>hotspots</i>	24
4.2	Pseudocódigo da função para segmentação de hotspots através de um grafo de ruas e esquinas.	26
4.3	Pseudocódigo da Poda de um grafo de hotspots para segmentação de componentes conexos com similaridade no comportamento temporal.	27
4.4	Análise de roubos na vizinhança de esquinas de Maceió/AL.	28
4.5	Segmentação de hotspots a partir da poda do grafo de Maceió/AL.	28
4.6	Série temporal de um <i>hotspot</i> , sendo decomposta em cada esquina.	28
5.1	Interface do Space-Time Visual Explorer composta por 4 componentes visuais: Index Range Slider (A), Crime and Patrolling Map View (B), Indexed Corners Scatterplot (C), Spatial-Time HeatMap (D)	30

6.1	Visão da distribuição geográfica dos crimes por vizinhança de cada esquina entre os anos de 2019 e 2021.	33
6.2	Seleção de grupamentos de esquinas a partir da quantidade de crimes ocorridos nas suas vizinhanças.	34
6.3	Componente visual <i>Index Scatterplot</i> da quantidade de crimes numa esquina (Y) pelo respectivo índice de ordenação de Fiedler no grafo (X), apresentado junto com o <i>Crime and Patrol Map View</i> com a dimensão geográfica selecionada e colorida de acordo com os crimes de vizinhança das unidades.	34
6.4	<i>Index Scatterplot</i> demonstrando distribuições de crimes em vizinhança por esquina, com filtro aplicado no <i>Crime and Patrol Map View</i>	35
6.5	Agregações trimestrais por ano do <i>Spatial-Time Heatmap</i> para visualização de série de temporal de um hotspot.	35
6.6	Distribuição global de passagens de viaturas em Maceió nos Trimestres de 1 a 3 durante o ano de 2021.	37
6.7	Distribuição global de passagens de viaturas em Maceió nos Trimestres de 1 a 3 durante o ano de 2021 após a filtragem que exclui os hotspots.	37
6.8	Distribuição espacial e temporal do patrulhamento na microrregião durante os três primeiros trimestres de 2021, agregadas por mês.	38

Lista de Tabelas

5.1	Tarefas Analíticas x Componentes Visuais.	30
-----	---	----

Lista de Abreviaturas e Siglas

IA	Inteligência Artificial
PMAL	Polícia Militar do Estado de Alagoas
PCA	Principal Component Analysis
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
KDE	Kernel Density Estimation
MSKDE	Marching Squares Kernel Density Estimation
TSNE	T-distributed Stochastic Neighbor Embedding

Sumário

Lista de Abreviaturas e Siglas	ix
1 Introdução	1
1.1 Objetivos	2
1.1.1 Objetivo Geral	3
1.1.2 Objetivos Específicos	3
2 Referencial Teórico	5
2.1 Estruturas de Dados Espaciais	5
2.1.1 Kd-Tree	6
2.1.2 Fórmula de Haversine e Ball-Tree	6
2.1.3 Subdivisão de Segmentos de Reta	9
2.1.4 Análise de dados Geoespaciais com grafos	11
2.1.5 Ordenação de Fiedler	12
2.2 Séries Temporais	15
3 Trabalhos Relacionados	18
3.1 Identificação de Padrões Criminais	19
3.2 Técnicas para Identificação de <i>Hotspots</i>	19
3.3 Viaturas e Patrulhamento Ostensivo	20
4 Metodologia	22
4.1 Bases de dados	23
4.2 Preprocessamentos e integração	23
4.3 Detecção de hotspots usando heurística de poda nos nós de um grafo	25
5 Space-Time Visual Explorer	29
5.1 Componentes visuais	30

6	Resultados	32
6.1	Contextualização	32
6.2	Estudo de Caso	33
7	Conclusão e Trabalhos Futuros	39
	Referências bibliográficas	41

Capítulo 1

Introdução

Nos últimos anos, temos observado uma crescente ênfase nos dados e nas inúmeras possibilidades de gerenciamento de processos e resultados, acompanhada da promoção de uma cultura organizacional orientada por dados, também conhecida como *Data Driven Culture*. Conforme observado por [Vargas, Alves e Mrejen \(2021\)](#), a pandemia da COVID-19 acelerou significativamente os processos tecnológicos, principalmente devido à necessidade de manter o distanciamento social, levando as organizações a buscar alternativas para manter suas operações. No entanto, essa mudança em grande escala também evidenciou que muitas empresas e organizações não estavam preparadas para enfrentar essa nova realidade, uma vez que não possuíam uma cultura de gestão e tomada de decisões totalmente baseada em dados.

Nesse contexto, agências de segurança pública e a comunidade científica têm se focado em desenvolver ferramentas inovadoras de Ciência de Dados com o intuito de analisar dados e reconhecer padrões; além de levantar e testar hipóteses usando dados. Dentro deste aspecto, no Brasil, um exemplo relevante do uso de ferramentas e técnicas de ciência de dados é o Núcleo de Estudos da Violência da USP (NEV-USP). Conforme observado por [Tonry, Ghirotti e Morris \(2003\)](#), é crucial que as forças de segurança pública adquiram conhecimento para embasar suas decisões e acompanhar a evolução da dinâmica social, bem como o agravamento das questões relacionadas à violência e segurança.

Historicamente, a criminologia comprova que a análise da evolução temporal dos crimes é fundamental para compreender e antecipar os locais onde as ocorrências criminais tendem a se concentrar ([Camacho-Collados e Liberatore, 2015](#)). Uma das estratégias para reduzir a incidência de crimes é o patrulhamento ostensivo e preventivo. No entanto, os métodos atuais de patrulhamento carecem de estudos mais aprofundados sobre sua eficácia ([Leigh, Dunnnett e Jackson, 2019](#)). Portanto, torna-se imperativo o uso eficaz de dados provenientes de diversas fontes, a fim de possibilitar a implementação de ações e políticas que visem à otimização dos recursos policiais, à repressão de crimes e à melhoria da qualidade do atendimento à população.

Porém, analisar grandes volumes de dados espaço-temporais é uma tarefa complexa. Em geral, ferramentas de visualização consideram apenas espaços visuais bidimensionais, visto que a visão humana é inerentemente bidimensional. Porém, dados urbanos georreferenciados são representados espacialmente por pares de coordenadas no formato latitude-longitude. Por outro lado, a dimensão temporal requer pelo menos mais 1 dimensão. Isto torna a tarefa de mapear estes dados para um espaço visual bidimensional desafiadora.

Em Alagoas, a Polícia Militar do Estado de Alagoas (PMAL) já mantém sistemas de informação que coletam e armazenam informações espaciais e temporais, por exemplo, de ocorrências criminais e das rotas de viaturas. Contudo, até mesmo em razão do volume de dados, técnicas específicas são necessárias para organizar, limpar e extrair conhecimento dessa massa. A grande dificuldade observada pela PMAL é justamente como processar esses dados e estabelecer formas de visualização que tragam significado para que possam usar de maneira efetiva a informação em favor da tomada de decisão orientada, de fato, aos dados que estão armazenados pela organização. Neste trabalho, usaremos bases espaço-temporais de dados criminais e de patrulhamento de viaturas da PMAL como estudo de caso para validar ferramentas capazes de analisar dados urbanos espaço-temporais e detectar manchas criminais de forma interativa.

Este trabalho apresenta ferramentas de análise visual capazes de relevar padrões em grandes volumes de dados urbanos espaço-temporais. Serão apresentados também preprocessamentos relevantes capazes de gerar representações concisas e computacionalmente eficientes destes grandes volumes de dados. Restringimos o escopo deste trabalho a dados espaço-temporais descrevendo ocorrências criminais e patrulhamento com viaturas. Estas representações devem viabilizar a aplicação de algoritmos de análise de dados, bem como a implementação de ferramentas de análise visual onde seja possível analisar a atividade criminal e o patrulhamento no espaço e no tempo. Nos estudos de caso apresentados, serão revelados padrões relevantes da relação entre a ocorrência de crimes e o patrulhamento.

1.1 Objetivos

Este capítulo desempenha um papel fundamental na definição e delimitação dos objetivos da presente pesquisa. Os objetivos estabelecidos fornecem a estrutura necessária para orientar o desenvolvimento e a execução deste estudo de mestrado em computação, garantindo sua coerência, relevância e contribuição para o avanço do conhecimento nesta área específica.

Os objetivos aqui delineados estão divididos em duas seções distintas: objetivo geral e objetivos específicos. O objetivo geral oferece uma visão ampla e abrangente dos propósitos principais desta pesquisa, delineando suas aspirações mais amplas e os resultados esperados. Por outro lado, os objetivos específicos detalham as metas mais concretas e os passos específicos a

serem seguidos para alcançar os objetivos gerais estabelecidos.

Essas seções trabalham em conjunto para fornecer uma estrutura clara e orientadora para o desenvolvimento deste estudo, assegurando que cada etapa da pesquisa contribua de maneira significativa para o alcance dos objetivos propostos. A definição cuidadosa desses objetivos é essencial para orientar o escopo da investigação, garantindo sua relevância acadêmica e aplicabilidade prática.

1.1.1 Objetivo Geral

O principal objetivo desta pesquisa é o desenho de inovadoras ferramentas de análise visual capazes de revelar padrões espaço-temporais relevantes em grandes volumes de dados urbanos espaço-temporais referentes a ocorrências criminais e patrulhamento de viaturas.

O principal propósito desta pesquisa é desenvolver ferramentas de análise visual inovadoras, destinadas a revelar padrões espaço-temporais significativos em grandes conjuntos de dados urbanos relacionados a ocorrências criminais e atividades de patrulhamento de viaturas. Essas ferramentas serão projetadas para revelar padrões valiosos sobre a dinâmica espacial e temporal desses eventos, permitindo uma compreensão mais profunda dos padrões de criminalidade e das estratégias de policiamento em ambientes urbanos.

1.1.2 Objetivos Específicos

Os objetivos específicos desta dissertação de mestrado são delineados para guiar e direcionar o desenvolvimento da pesquisa de forma clara e precisa.

Inicialmente, será realizada uma análise exploratória dos dados, seguida, se necessário, pela aplicação de medidas de sanitização e pela criação de novos atributos. Essas etapas visam ampliar o valor informativo dos dados, preparando-os para a aplicação de técnicas de reconhecimento de padrões.

Em seguida, será conduzida uma pesquisa aprofundada para o desenvolvimento de algoritmos customizados capazes de lidar eficazmente com o processamento de grandes volumes de dados. Isso será crucial devido à magnitude dos dados provenientes das ocorrências criminais e das atividades de patrulhamento, bem como à necessidade de análises em larga escala.

Posteriormente, serão investigadas e aplicadas técnicas de análise visual para compreender os padrões espaço-temporais das ocorrências criminais e de patrulhamento de viaturas, assim como padrões relacionados a outros atributos relevantes.

Além disso, serão exploradas metodologias inovadoras para identificação e processamento de dados de hotspots (manchas criminais), visando uma abordagem mais eficiente e

precisa na análise dessas áreas.

A pesquisa também incluirá um estudo de caso abrangente, utilizando as bases de dados fornecidas pela PMAL, culminando na elaboração de um artefato de software destinado à análise das informações obtidas.

Por fim, será identificada a correlação entre as ocorrências criminais e as atividades de patrulhamento das viaturas, utilizando as ferramentas desenvolvidas ao longo da pesquisa. Essa análise será fundamental para compreender os padrões de comportamento criminal e para o aprimoramento das estratégias de policiamento preventivo.

Capítulo 2

Referencial Teórico

Esta seção explora conceitos fundamentais relacionados à análise e processamento de dados geoespaciais. Inicialmente, apresentaremos a importância das Estruturas de Dados Espaciais na representação e organização de informações geográficas. Em seguida, discutiremos a Kd-Tree, uma estrutura amplamente utilizada na resolução de problemas de busca espacial. Abordaremos a Fórmula de Haversine e a Ball-Tree, focando na aplicação dessa métrica para cálculos de distâncias geoespaciais e na construção de árvores otimizadas para pesquisa. Continuaremos com a exploração da Análise de Dados Geoespaciais em Grafos, apresentando conceitos-chave e sua aplicação na compreensão de relacionamentos e fluxos em ambientes geográficos. Abordaremos também a técnica de Ordenação de Fiedler, uma abordagem essencial para a detecção de comunidades em redes geoespaciais. Finalmente, abordaremos a análise de Séries Temporais, explorando como essa técnica é utilizada na identificação de padrões temporais em eventos geoespaciais e no contexto das aplicações de segurança pública.

2.1 Estruturas de Dados Espaciais

Problemas envolvendo a análise de dados espaciais em larga escala demandam o uso de estruturas de dados específicas para realizar buscas eficientes. Na análise de dados espaciais, uma das métricas essenciais é a distância entre pontos na superfície da Terra, que equivale à distância entre pontos em uma esfera.

Em situações mais simples, como cálculos de distância em um plano bidimensional, a geometria euclidiana tradicional é adequada e estruturas de dados mais simples podem ser empregadas para buscas eficazes. No entanto, quando lidamos com análises geoespaciais, essa abordagem não se mostra apropriada. As coordenadas cartográficas referenciam a posição com base na latitude e longitude, medidas em relação ao equador e ao meridiano de Greenwich. Um ponto na Terra é localizado por meio de um par de coordenadas: latitude e longitude. A

latitude de um ponto é o ângulo entre o plano equatorial e a linha que passa pelo centro da Terra e por esse ponto, enquanto a longitude é o ângulo leste ou oeste em relação a um meridiano de referência que passa por esse ponto. A combinação de latitude e longitude é a entrada para métricas espaciais na esfera, como Haversine, que adotamos neste trabalho.

Dado o exposto, algumas estruturas de dados podem não ser adequadas para lidar com análises espaciais desse tipo. Esse é o caso da *Kd-Tree*, uma estrutura de árvore amplamente usada em problemas de busca espacial, mas inadequada para métricas relacionadas ao espaço esférico. Para tais problemas, existem outras estruturas, como a *Ball-Tree*, que possibilitam o uso de métricas adequadas, como a distância de Haversine, ao mesmo tempo em que oferecem eficiência em buscas. Descreveremos ambas a seguir.

2.1.1 Kd-Tree

Uma das estruturas de dados comumente utilizadas na organização de dados em espaços euclidianos de k dimensões é a *Kd-Tree*. Conforme mencionado por Berg et al. (r, e), essa estrutura de árvore utiliza a mediana das coordenadas para subdividir conjuntos de pontos, criando uma estrutura organizada para percorrer a árvore. A divisão ocorre alternadamente entre cada eixo, e cada sub-região é subdividida através da mediana das coordenadas dos pontos, como demonstrado no pseudocódigo do algoritmo na Figura 2.1.

Nesses casos, estruturas de dados como a *Ball-Tree* se tornam mais adequadas, pois permitem a aplicação de métricas personalizadas, como a Fórmula de Haversine, ao mesmo tempo em que mantêm uma eficiência de busca. A *Ball-Tree* é capaz de organizar dados geoespaciais de maneira eficiente, levando em consideração a curvatura da Terra e a verdadeira distância entre os pontos.

Portanto, ao lidar com problemas de análise de dados geoespaciais, é essencial escolher a estrutura de dados apropriada, como a *Ball-Tree*, que pode lidar com métricas de distância não euclidianas e garantir um desempenho eficiente na busca.

2.1.2 Fórmula de Haversine e Ball-Tree

A Fórmula de Haversine deriva da Lei Esférica dos Cossenos DAUNI et al., 2019 Ela é utilizada para calcular a menor distância entre dois pontos em uma esfera, seguindo o círculo máximo que passa por esses pontos. A fórmula leva em consideração as coordenadas de latitude e longitude dos dois pontos, conforme apresentado na Equação (2.1), onde lat_1 é a latitude do primeiro ponto, lon_1 é a longitude do primeiro ponto, lat_2 é a latitude do segundo ponto e lon_2 é a longitude do segundo ponto. A fórmula calcula a distância entre esses dois pontos na esfera em unidades de radianos.

A Fórmula de Haversine é particularmente relevante em problemas de análise de dados geoespaciais, onde a distância real entre pontos na superfície da Terra precisa ser calculada considerando a curvatura da esfera. Essa fórmula é essencial para o cálculo de distâncias geoespaciais em estruturas de dados como a *Ball-Tree*, que permitem a organização eficiente de

Algoritmo 1: Função Kd-Tree

Entrada: Um conjunto de pontos P e a profundidade atual $prof$
Saída: A raiz de uma árvore k-D armazenando P
 ConstroiArvKDT(P , $prof$)
 se P contém apenas um ponto então
 Saída: u
 ma folha contendo este ponto
 fim
 senão
 se $prof // \text{número de pontos} = 0$ então
 divida P em dois subconjuntos com uma linha vertical pela coordenada x
 mediana dos pontos em P ;
 P_1 é o conjunto de pontos à esquerda da linha;
 Os pontos exatamente na linha pertencem à P_1 ;
 P_2 o conjunto de pontos da direita;
 fim
 senão
 divida P em dois subconjuntos com uma linha horizontal pela coordenada y
 mediana dos pontos em P ;
 P_1 é o conjunto de pontos acima da linha;
 Os pontos exatamente na linha pertencem à P_1 ;
 P_2 o conjunto de pontos abaixo da linha;
 fim
 fim
 $V_{dir} \leftarrow \text{ConstroiArvKDT}(P_1, prof + 1)$
 $V_{esq} \leftarrow \text{ConstroiArvKDT}(P_2, prof + 1)$
 Crie um nó V ;
 Atribua V_{dir} e V_{esq} juntamente como filhos direito e esquerdo de v , respectivamente;
Saída: V

Figura 2.1: Pseudocódigo da *KD-Tree*

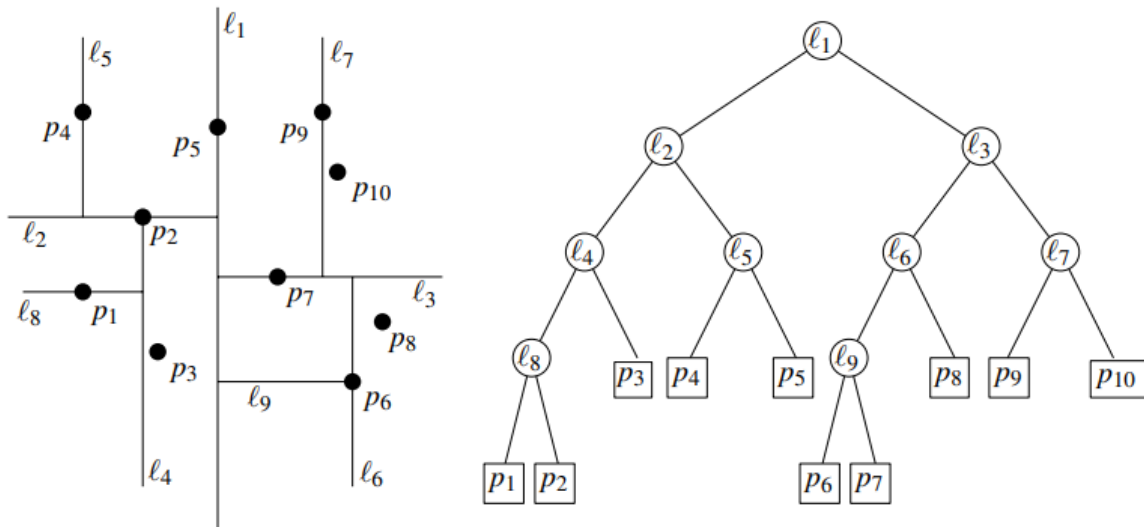


Figura 2.2: Exemplo de estrutura de organização dos nós de uma *Kd-Tree*.

dados geográficos.

$$D(lat_1, lon_1, lat_2, lon_2) = 111,17 \cdot (\arccos(\sin(lat_2) \sin(lat_1) + \cos(lat_2) \cos(lat_1) \cos(lon_2 - lon_1))) \quad (2.1)$$

Nesse contexto, uma variação da estrutura de dados *Kd-Tree*, conhecida como *Ball-Tree*, apresenta-se como uma alternativa adequada para representar pontos em um espaço métrico, levando em consideração a distância calculada pela Fórmula de Haversine. A *Ball-Tree* é um método baseado na estratégia de divisão e conquista, onde o nó é dividido em duas partições sucessivamente. A construção de uma *Ball-Tree* pode ser realizada de acordo com o pseudocódigo apresentado no Algoritmo 2.3. É importante ressaltar que a distância utilizada nesse algoritmo pode ser definida como a distância de Haversine. A *Ball-Tree* oferece a capacidade de representar eficazmente dados geoespaciais, tornando-se uma escolha adequada para resolver problemas de análise espacial que envolvem Haversine como métrica de distância. Essa estrutura de dados é particularmente relevante para a organização eficiente de dados geográficos, permitindo a realização de consultas de maneira eficaz em problemas de Ciência de Dados que envolvem dados geoespaciais.

Na Figura 2.4, a linha verde ilustra o hiperplano utilizado para dividir um conjunto de

Algoritmo 2: Função Constrói Ball-Tree

Data: P, Um conjunto de pontos
Entrada: B, A raiz de uma árvore Ball-Tree armazenando P
 ConstroiArvBLT(P)
 se P contém apenas um ponto então
 Saída: U
 ma folha contendo este ponto
 fim
 senão
 seja c a dimensão de maior dispersão;
 seja k a mediana da dimensão c;
 seja p o ponto central selecionado considerando c;
 seja L conjuntos de pontos à esquerda de k;
 seja R conjuntos de pontos à direita de k;
 crie B com dois filhos:
 B.pivot ← p;
 B.child1 ← ConstroiArvBLT(L);
 B.child2 ← ConstroiArvBLT(R);
 seja B.raio a distância máxima de p entre os filhos;
 fim
Saída: B

Figura 2.3: Pseudocódigo de construção da *Balltree*.

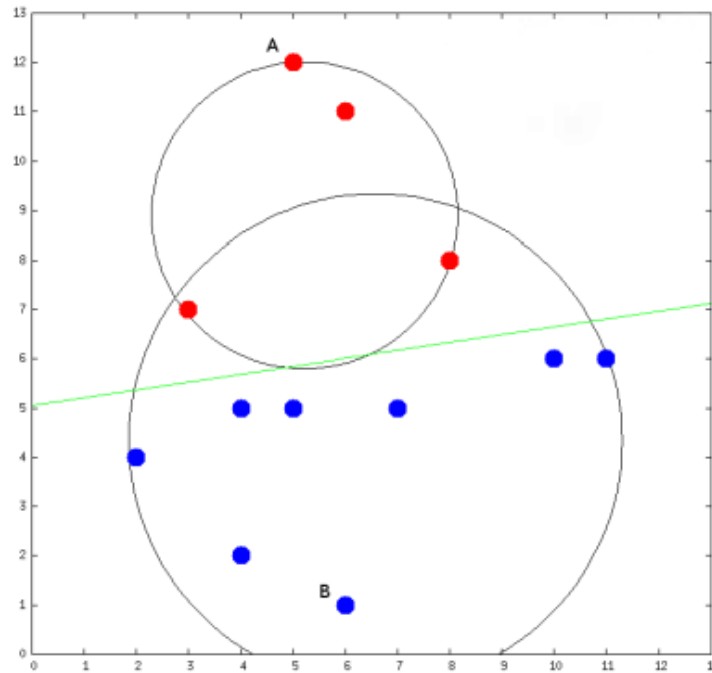


Figura 2.4: Primeiro particionamento de pontos em um plano definidos numa *Ball-Tree*.

pontos em dois subconjuntos. Essa divisão resulta em nós filhos na estrutura da *ball-tree*, os quais representam os subconjuntos derivados da partição do conjunto de pontos original. Conforme destacado por [DOLATSHAH; HADIAN; MINAEI-BIDGOLI, 2015](#) a divisão baseada nesse recurso pode, em alguns casos, levar à construção de uma árvore desbalanceada. Para mitigar esse problema e a limitação associada à escolha do hiperplano, a Análise de Componentes Principais (PCA) é empregada. A PCA permite encontrar a direção de maior dispersão nos dados, representada pelo autovetor associado ao maior autovalor da matriz de covariância dos dados. O hiperplano perpendicular a essa direção é então selecionado para a divisão dos dados.

Além disso, o centróide, que é o ponto representado pelas coordenadas médias de todos os pontos no conjunto, é utilizado como referência para a escolha do hiperplano. Essa abordagem contribui para uma melhor organização da *ball-tree* e, por conseguinte, otimiza a representação e busca de dados geoespaciais. O mecanismo de consulta da *Ball-Tree* traz uma metodologia baseada em uma busca em profundidade, como exibido no algoritmo representado na Figura 2.5.

2.1.3 Subdivisão de Segmentos de Reta

Nesta seção descrevemos como subdividir um segmento de reta. Este algoritmo de Geometria Computacional é essencial para realizar a discretização espacial dos crimes usando as esquinas como unidades espaciais.

Dado um espaço vetorial V , um subconjunto desse espaço R poderá ser considerado um segmento de reta se pode ser representado de forma paramétrica através da Equação (2.2), onde

u e $u + v$ representam as extremidades do segmento de reta, tal que $u, v \in V$.

$$R(t) = \{u + t \cdot v \mid t \in [0, 1]\} \quad (2.2)$$

A subdivisão de um segmento de reta pode ser realizada de maneira uniforme, amostrando-se o segmento R uniformemente o parâmetro t no intervalo $[0, 1]$. Mais especificamente, considerando n a quantidade de pontos desejada, faça $\Delta = \frac{1}{n-1}$. Daí

$$\{R(u + i \cdot \Delta v) \mid i = 0 \dots n-1\}$$

fornece uma amostra uniforme do segmento.

A subdivisão de segmentos de reta será usada neste trabalho para subdividir o grafo de ruas (representação da estrutura espacial de um local, onde vértices são esquinas e arestas são ruas), permitindo a realização de buscas eficientes entre localizações da cidade e esquinas mais próximas, sem a necessidade de usar algoritmos de busca entre ponto e segmentos de reta. De forma similar, [Garcia-Zanabria et al. \(2020\)](#) utilizam métodos para projetar locais (pontos) na rua mais próxima (aresta), permitindo que dados possam ser geograficamente discretizados e organizados para melhor visualização, como é possível observar na Figura 2.1.3. Mais especificamente,

Algoritmo 3: Busca Vizinho Mais Próximo

Entrada:

t - o ponto de destino para a consulta

k - o número de vizinhos mais próximos de t para procurar

Q - fila de prioridade máxima contendo no máximo k pontos

B - um nó, ou bola, na árvore

Saída: Q , contendo os k vizinhos mais próximos de dentro de B

BuscaKNN(t, k, Q, B)

se $distance(t, B.pivotal) - B.raio \leq distancia(t, Q.primeiro)$ **então**

 | **Saída:** Q *inalterado*

fim

senão

se B *for um nó folha* **então**

para cada ponto p *em* B **faça**

se $distancia(t, p) < distancia(t, Q.primeiro)$ **então**

 | *adicione* p *a* Q **se** $se tamanho(Q) > k$ **então**

 | *remova o vizinho mais distante de* Q

fim

fim

seja $filho1$ *o nó filho mais próximo de* t

seja $filho2$ *o nó filho mais distante de* t

 BuscaKNN($t, k, Q, filho1$)

 BuscaKNN($t, k, Q, filho2$)

fim

fim

fim

Saída: Q

Figura 2.5: Pseudocódigo da busca de nós vizinhos na estrutura de *Balltree*.

pontos foram associadas aos vértices mais próximos (esquinas), não considerando a distância euclidiana, mas a distância baseada na projeção do ponto junto à aresta (rua) mais próxima.

Diferentemente, pretende-se neste trabalho usar a subdivisão dos segmentos de retas para transformar o problema de busca de aresta mais próxima, em um problema (aproximado) de busca de ponto mais próximo (os pontos das arestas subdivididas), como exibido na Figura 2.7. As arestas das ruas são subdivididas em uma série de pontos justapostos e equidistantes, que serão usados para representá-las. Assim, é possível encontrar o ponto u_n no espaço vetorial V que mais se aproxima do grupo de pontos e, em seguida, definir qual a esquina mais próxima de u_n , considerando o dado espaço vetorial.

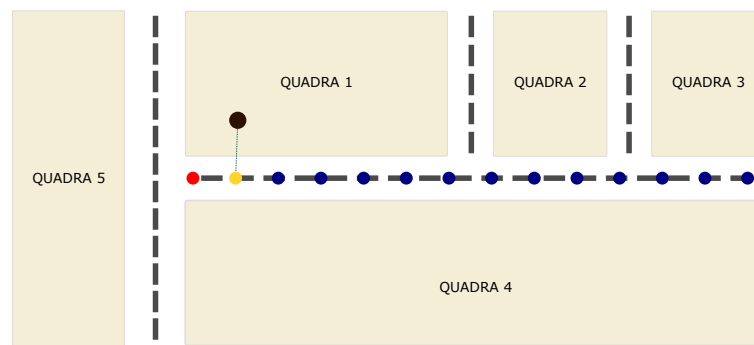


Figura 2.7: Subdivisão de uma rua (segmento de reta), de maneira a localizar a esquina mais próxima (ponto vermelho), a partir do ponto mais próximo da ocorrência no espaço vetorial V indicado pelos pontos. O ponto amarelo indica o ponto no espaço vetorial V mais próximo do ponto da ocorrência.

2.1.4 Análise de dados Geoespaciais com grafos

Para definição de um grafo, segundo [Rosen e Krithivasan \(2012\)](#), tem-se $G = (V, E)$, tal que G seja uma estrutura discreta que consiste num conjunto não vazio de nós, ou vértices, V e um grupo de arestas E , tal que essas arestas conectam dois vértices, também chamadas de extremidades.

Grafos podem ser não direcionados, quando suas arestas não possuem um sentido vinculado; ou podem ser direcionados, quando as arestas possuem um sentido específico para percorrer o grafo. Um grafo direcionado $G = (V, E)$, também chamado digrafo, consiste em um conjunto não vazio de vértices V e um conjunto de arestas direcionadas E (ou arcos). Cada aresta direcionada está associada a um par ordenado de vértices. Diz-se que a aresta direcionada associada ao par ordenado (u, v) começa em u e termina em v .

Existem algumas formas para representar estruturas de grafos, dentre elas, a representação através da matriz de adjacências ([Koshy, 2004](#)). O principal objetivo desta representação é fornecer um modelo algébrico capaz de correlacionar os elementos do conjunto de vértices V , num dado grafo G . É possível definir a matriz de adjacência A pela relação onde $A = [a_{ij}]$,

tal que $a_{ij} = 1$, sempre que houver uma aresta (v_i, v_j) pertencente ao grafo G , do contrário assume-se que $a_{ij} = 0$. Através da Figura 2.8 é possível estabelecer a matriz de adjacência do grafo demonstrado.

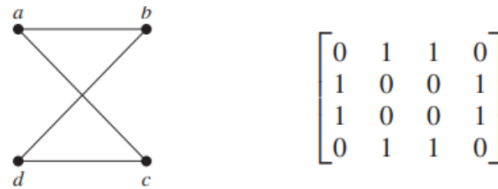


Figura 2.8: Representação da matriz de adjacência de um grafo.

Um dos problemas mais recorrentes que se busca solução dentro do escopo dos grafos é o de encontrar o caminho de menor custo que conecta dois vértices. O algoritmo de Dijkstra é conhecido justamente por resolver boa parte dos problemas que envolvem a busca de caminhos com menor peso entre dois nós num grafo, observando-se que os pesos atribuídos entre os vértices são não negativos. Inicialmente, o algoritmo ajusta todas as distâncias, exceto do vértice de origem, para indefinido ou infinito, com o intuito de inserir na memória de trabalho se um vértice já foi ou não visitado, como é possível observar no Algoritmo 2.9. Na próxima etapa do algoritmo, são atualizadas as distâncias entre os vértices que estão diretamente ligados ao vértice de origem, retira-se o vértice da lista dos não visitados e marca-o como visitado. Repetido o processo para os vértices não visitados até que o nó de destino seja marcado como visitado ou se a menor distância tentativa entre os nós do conjunto não visitado for infinita (quando não há conexão entre os nós restantes).

Considerando um peso atribuído a cada aresta do grafo, essa busca pode assumir uma outra metodologia, por exemplo, retornando uma lista de vértices que estão num determinado limite de distância do vértice de origem dentro da malha de grafos. Essa metodologia é essencialmente interessante para problemas de análise espacial onde se deseja aferir dados de uma dada vizinhança num vértice.

2.1.5 Ordenação de Fiedler

Na Teoria Espectral de Grafos, um conceito fundamental é o vetor de Fiedler. O vetor de Fiedler é derivado do segundo menor autovalor não nulo de uma matriz Laplaciana, que representa as propriedades estruturais de um grafo. Esse vetor desempenha um papel crucial em diversas aplicações, como particionamento de grafos e detecção de vizinhanças.

A ordenação de Fiedler é um processo que utiliza o vetor de Fiedler para reorganizar os vértices de um grafo de forma a revelar padrões estruturais e melhorar a visualização ou o desempenho de algoritmos. Além disso, a ordenação de Fiedler pode ser aplicada em algoritmos de particionamento de grafos. Ao ordenar os vértices de acordo com o vetor de Fiedler, é possível

dividir o grafo em subgrafos que sejam coesos e preservem a vizinhança entre os nós do grafo, com isso nós do grafo podem ser expressos num vetor unidimensional de acordo com os valores no vetor de Fiedler.

A ordenação de Fiedler tem várias aplicações práticas. Em visualização de grafos, por exemplo, a ordenação pode ser usada para posicionar os vértices de maneira que revele a estrutura subjacente do grafo de maneira mais clara e compreensível. Isso é particularmente útil quando se lida com grafos grandes e complexos, nos quais a visualização pode se tornar confusa sem uma organização adequada.

Uma matriz de adjacência é uma representação de um grafo através de uma matriz bidimensional, onde cada linha e coluna representa um vértice do grafo. Os elementos da matriz indicam se existe uma aresta entre os vértices correspondentes. Essa matriz, chamemos de A .

Uma matriz de graus de um grafo é uma representação matricial que descreve a conectividade dos vértices no grafo. Em uma matriz de graus, cada linha ou coluna corresponde a um vértice do grafo, e os elementos da matriz indicam o grau de cada vértice, ou seja, o número de arestas incidentes a esse vértice. Essa matriz, chamaremos de D .

Para compreender o processo de ordenação faz-se necessário compreender o Vetor de Fiedler. Considerando um grafo não direcionado G , sua matriz laplaciana L pode ser obtida

Algoritmo 4: Algoritmo de Dijkstra

Entrada: Grafo G , Vértice Origem

Saída: Distancia, Anterior

Dijkstra(G , Origem)

distancia[origem] = 0

Cria conjunto de vertices Não-Visitados

para cada vertice $\in G$ **faça**

se $v \neq \text{origem}$ **então**

 | distancia[v] = ∞

fim

fim

anterior[v] = Indefinido

Insere v em Não-Visitados

enquanto Não-Visitados não for vazio **faça**

Procura vértice com menor distância denominado u

Remove u de Não-Visitados

i vizinho de u

 custo = distancia[u] + $g[u, i]$

se custo < distancia[i] **então**

 | distancia[i] = custo

 | anterior[i] = u

fim

fim

Saída: distancia, anterior

Figura 2.9: Representação do Pseudocódigo de Dijkstra.

através da diferença entre a matriz de graus D e a matriz de adjacência A , observando-se que $L = D - A$.

Ao calcular a matriz laplaciana L do grafo G , encontra-se os autovetores $\vec{v}_n$ e autovalores λ_n da matriz L e, em sequência, busca-se a posição do segundo menor autovalor λ_2 na lista de autovalores. Através de λ_2 , obtendo-se o autovetor correspondente, é realizada uma ordenação dos nós do grafo G de acordo com o autovetor $\vec{v}_n$, retornando uma lista com a ordenação dos nós v .

Algoritmo 5: Vetor de Fiedler

Entrada: Grafo G
Saída: Lista ordenada de nós v
 $A \rightarrow \text{CalcularMatrizAdjacencias}(G)$
 $D \rightarrow \text{CalcularMatrizGraus}(G)$
 $L = D - A$
 $\lambda_n, \vec{v}_n = \text{CalculaAutoValoresAutoVetores}(L)$
 $\text{fiedlerAutovector} \rightarrow \vec{v}_n[\lambda_2]$
 $v \rightarrow \text{sortGraphNodes}(G, \text{fiedlerAutovector})$
Saída: v

Figura 2.10: Pseudocódigo de ordenação de um vetor de Fiedler.

A lista ordenada constitui um vetor unidimensional capaz de proporcionar em cada um de seus índices uma relação com um respectivo vértice do grafo G , onde posições próximas no vetor refletem nós também numa dada vizinhança, permitindo assim, uma partição coesa do grafo em regiões menores através de subgrafos.

Uma vez aplicada a ordenação de Fiedler, dados dois nós i e j em condição de adjacência, os respectivos valores em v revelam uma similaridade entre os vértices i e j no grafo. Vale ressaltar que os componentes de v são determinados pela minimização da equação ((2.3)), onde u_i e u_j representam os componentes do vetor u , que é um vetor de dimensão n , onde n é o número de vértices no grafo.

$$v = \min_{\|u\|=1} \sum_{\substack{\forall i \neq j s.t. \\ L_{ij} \neq 0}} |L_{ij}| (u_i - u_j)^2 \quad (2.3)$$

Uma vez ordenado o grafo com base em seu respectivo Vetor de Fiedler, será possível realizar análises de vizinhanças, agrupando vértices com um certo grau de similaridade espacial. Considerando um grafo que contém as esquinas da cidade de Maceió/AL, na Figura 2.11, é possível perceber a transição de cores no mapa onde cores mais similares denotam regiões mais próximas geograficamente, demonstrando que o índice de fiedler é capaz de preservar vizinhanças espaciais num espaço unidimensional, proporcionando recursos viáveis para a exploração visual de dados espaciais.

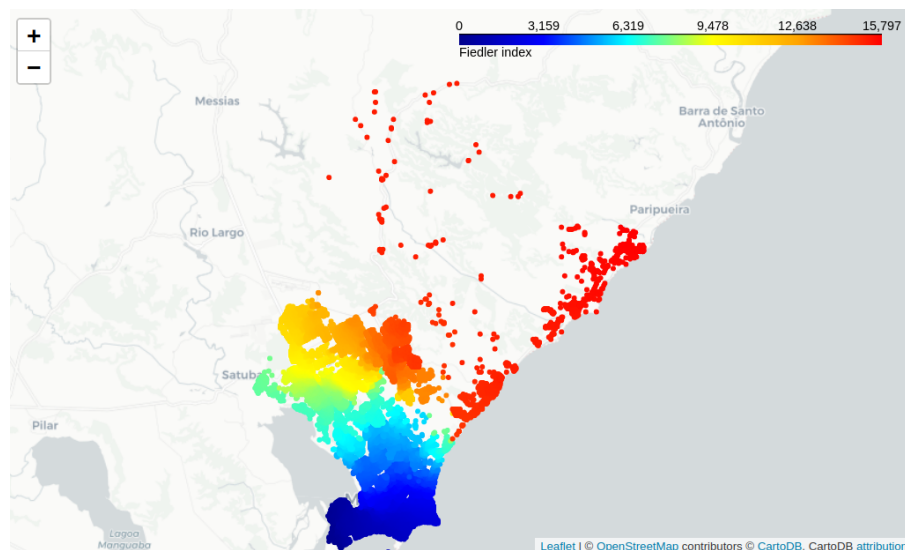


Figura 2.11: Grafo de Esquinas de Maceió/AL ordenado segundo o vetor de Fiedler.

2.2 Séries Temporais

Uma série temporal é um modelo matemático composto por um conjunto sequencial de observações de uma variável específica, organizadas de forma sequencial ao longo do tempo. Conforme discutido por [Ehlers \(2007\)](#), essas observações podem ser realizadas de maneira contínua ou de forma discreta, em intervalos específicos.

Uma série temporal pode ser decomposta em componentes de natureza determinística, passíveis de modelagem por meio de equações matemáticas. Contudo, também existem componentes estocásticos na série temporal, sujeitos ao acaso e dependentes do estado anterior da série. Esses componentes, interligados, desempenham papéis distintos na análise de séries temporais:

- **Tendência:** Captura elementos de longo prazo na série, podendo ser modelada por equações matemáticas ou por meio de componentes estocásticos.
- **Ciclo:** Representa variações mais ou menos longas em torno de uma tendência estabelecida.
- **Sazonalidade:** Estima padrões que envolvem ciclos regulares específicos ao longo do tempo, adicionando uma dimensão previsível às observações.
- **Aleatoriedade:** Em conjunto com os elementos anteriores, tem a função de capturar efeitos não observados nos componentes determinísticos e sazonais

A observação do comportamento de variáveis ao longo do tempo, quando combinada com a análise espacial, oferece insights valiosos para validar ou refutar hipóteses ao explorar as interações entre essas variáveis nas dimensões espaço e tempo. No contexto do estudo de padrões criminais, conforme enfatizado por [Garcia-ZANABRIA et al. \(2021\)](#), a análise temporal

desempenha um papel crucial na geração de atributos que facilitam a identificação e estudo de *hotspots* com comportamentos consistentes ao longo do tempo. Isso inclui a descoberta de padrões sazonais, contribuindo para a coordenação eficiente das operações das forças de segurança pública.

A Figura 2.12 ilustra a metodologia de visualização do Cripav (Garcia-ZANABRIA et al., 2021), uma ferramenta analítica de dados espaciais e temporais de crime. A visualização relaciona padrões a elementos de infraestrutura, evidenciando mudanças ao longo do tempo. Essa abordagem integrada demonstra a eficácia na compreensão aprofundada de fenômenos complexos, enfatizando a interconexão entre os domínios espacial e temporal na análise de séries temporais.

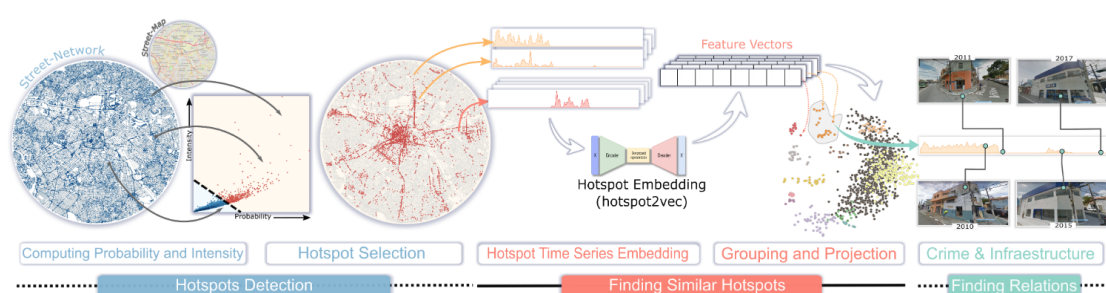


Figura 2.12: Metodologia de visualização do Cripav para análise espacial e temporal de crimes, relacionando esses padrões à infraestrutura e sua mudança no tempo (Garcia-ZANABRIA et al., 2021).

A estruturação adequada de dados para análises temporais é um aspecto crucial em pesquisas que buscam compreender padrões e dinâmicas ao longo do tempo. Inicialmente, é imperativo adotar uma representação temporal precisa, utilizando formatos de data e hora padronizados para garantir a consistência nas observações. A escolha da granularidade temporal também é uma consideração essencial, variando de segundos a anos conforme os objetivos da pesquisa.

Ao desenvolver uma amostragem temporal equitativa, evita-se viés nos resultados, sendo fundamental selecionar intervalos regulares e significativos. Além disso, é necessário lidar estrategicamente com dados temporais ausentes, seja por meio de interpolação, imputação ou ao considerar a ausência como uma informação válida.

Outro ponto importante é a possibilidade de aplicar transformações às variáveis para estabilizar a variância ao longo do tempo, tornando os dados mais propícios a análises estatísticas. A inclusão de variáveis temporais auxiliares, como feriados e eventos significativos, enriquece a contextualização temporal.

A resolução temporal deve ser criteriosamente escolhida, de acordo com os eventos de interesse, podendo ser necessário agregar ou desagregar os dados. A documentação detalhada de metadados sobre a coleta e processamento dos dados é crucial para transparência, replicabilidade

e validação dos resultados.

Por fim, a escolha de ferramentas e plataformas adequadas é fundamental, considerando a escala e complexidade do conjunto de dados temporais. Seguindo essas diretrizes científicas, os pesquisadores podem estruturar dados temporais de maneira robusta, estabelecendo uma base sólida para análises detalhadas e interpretações significativas ao longo do tempo.

Capítulo 3

Trabalhos Relacionados

Dentro do âmbito da segurança pública, um tópico de extrema relevância é a avaliação do impacto das ferramentas tecnológicas, tanto de hardware quanto de software, na administração da segurança e na qualidade de vida da população. De acordo com [Tonry, Ghirotti e Morris \(2003\)](#), essa preocupação representa uma diretriz fundamental para as pesquisas e esforços voltados para a resolução de uma ampla gama de desafios relacionados à segurança pública. Portanto, torna-se imperativo realizar uma análise minuciosa e abrangente dos estudos e pesquisas já conduzidos, a fim de identificar abordagens passíveis de incorporação e personalização, com o propósito de alcançar os objetivos estabelecidos nesta pesquisa.

Nesta seção, serão apresentadas de maneira sucinta algumas contribuições científicas relevantes ao escopo desta pesquisa, as quais desempenham um papel substancial no enriquecimento e orientação do seu desenvolvimento. A primeira abordagem diz respeito à Identificação de Padrões Criminais, seguida pela seção subsequente dedicada à Identificação de *Hotspots*. Por fim, a terceira seção abordará a temática de Viaturas e Patrulhamento Ostensivo.

Na seção destinada à Identificação de Padrões Criminais, serão discutidos resultados provenientes de pesquisas bibliográficas relacionadas ao tema, concentrando-se na aplicação de técnicas e modelos capazes de identificar padrões de atividade criminosa. A seção subsequente, que trata das Técnicas de Identificação de *Hotspots*, explorará diversas abordagens de pesquisa na área, incluindo uma análise comparativa que realçará aspectos críticos nas respectivas implementações, contribuindo, assim, significativamente para esta pesquisa. Na última seção, concernente a Viaturas e Patrulhamento Ostensivo, serão resumidos trabalhos relevantes que abordam aspectos metodológicos, estudos estatísticos e perspectivas sociológicas relacionados às implicações do policiamento ostensivo e suas variantes.

3.1 Identificação de Padrões Criminais

A identificação de padrões de crime exige notável complexidade na análise e interpretação de dados espaço-temporais, cuja dificuldade reside na identificação de similaridades nas linhas temporais de espaços previamente discretizados. Para abordar esse desafio, pesquisadores têm recorrido a uma variedade de algoritmos e técnicas, incluindo o *K-Means* (Garcia et al., 2021), *DBSCAN*, e até mesmo abordagens baseadas em *Deep Learning* (Garcia-ZANABRIA et al., 2021). Em termos de análise temporal, a investigação da sazonalidade dos eventos tem se revelado valiosa não apenas para compreender similaridades entre ocorrências, mas também para destacar padrões temporais, um aspecto de relevância significativa na análise de crimes (Garcia-ZANABRIA et al., 2021).

Trabalhos anteriores, como os estudos conduzidos por Garcia-Zanabria et al. (2020) e Garcia et al. (2021), identificaram certas limitações nas implementações de análise geográfica. Algumas dessas limitações se devem à adoção de medidas de distância baseadas na geometria euclidiana entre pontos, em vez de utilizar distâncias reais entre pontos na malha do grafo da cidade, como o estudo de caso de (Garcia-Zanabria et al., 2020), onde descreve sua ferramenta analítica *Mirante*. Outro desafio relaciona-se à utilização de ruas (arestas) como unidades de análise para concentrar informações de crimes, o que pode complicar a análise em áreas urbanas extensas, onde ruas de grande comprimento frequentemente apresentam características geográficas diversas, como é evidenciado por (Garcia-Zanabria et al., 2020).

O trabalho realizado por Garcia et al. (2021), denominado *Cripav*, merece destaque pela ênfase na identificação de *hotspots*. Esse estudo se baseia na agregação de dados de crimes nas esquinas da cidade. A abordagem adotada permitiu uma melhor identificação de pontos similares e a posterior identificação de *hotspots* com base em um equilíbrio entre intensidade e probabilidade, construído por meio de um processo Markoviano. Esse processo envolve a aplicação de dados de crimes para construir matrizes estocásticas.

3.2 Técnicas para Identificação de *Hotspots*

As técnicas voltadas para a identificação de *hotspots* têm como objetivo a localização de áreas discretas com alto potencial de ocorrências criminais. Um dos principais desafios na análise espacial de crimes, conforme ressaltado por Garcia et al. (2021), reside na compreensão e estabelecimento das unidades espaciais apropriadas para a análise. A escolha dessas unidades desempenha um papel crítico e pode influenciar significativamente os resultados da análise, sendo considerada por alguns como o objetivo central de análises de predição de crimes.

A detecção de *hotspots* abrange uma ampla variedade de técnicas documentadas na literatura. Diversas metodologias foram desenvolvidas para discretizar o espaço em análises

geográficas, incluindo o uso de malhas de grafos que utilizam as arestas para representar as ocorrências criminais (Garcia-Zanabria et al., 2020), a divisão do espaço em quadrantes com base em grades de latitude e longitude (Garcia et al., 2021), e esquinas (Garcia-ZANABRIA et al., 2021), como adotado neste trabalho. Uma das técnicas derivadas da discretização do espaço é a Estimção da Densidade de Núcleo, KDE (Bailey, Gatrell et al., 1995). Essa abordagem foi considerada promissora por Leigh, Dunnett e Jackson (2019), pois reduz os efeitos das fronteiras geográficas, levando em conta crimes nas áreas circundantes. O KDE é aplicado em uma grade sobreposada ao mapa de ruas e aos dados de crimes. Além disso, foram desenvolvidas variações do KDE, como o KDE de Quadrados Marchantes (*Marching Squares KDE*, MSKDE) (Neto José Florêncio e Santos,), o KDE Limitado por Rede (*Network-constrained KDE*, NKDE) (Nie et al., 2015), e o *Network-constrained Getis-Ord Gi* (Nie et al., 2015).

Outras técnicas para detectar *hotspots* incluem métodos baseados em pontos e glifos (Ratcliffe, 2010), mapeamento coroplético (Hirschfield e Bowers, 2001), autocorrelação (Craglia, Haining e Wiles; Ratcliffe e McCullagh, 2000, 1999), agrupamento (Anderson; Grubestic, 2009, 2006) e análise estatística espaço-temporal (Lodha e Verma, 2000). Para Garcia-ZANABRIA et al. (2021), os *hotspots* são detectados visualmente através de um gráfico de dispersão de Probabilidade $\times$ Intensidade, onde a probabilidade de crimes é estimada através de uma abordagem Markoviana. Assim como neste último trabalho, também propomos uma ferramenta visual para analisar e detectar *hotspots* de forma interativa, mas utilizando algoritmos de ordenação espacial de grafos.

3.3 Viaturas e Patrulhamento Ostensivo

A atividade de patrulhamento ostensivo é uma componente intrínseca da segurança pública, desempenhando um papel crucial na garantia da segurança da comunidade e na prevenção da atividade criminosa. Esta atividade se traduz na materialização do serviço de segurança, buscando coagir a criminalidade e manter o equilíbrio social. Camacho-Collados e Liberatore (2015) propuseram um sistema de suporte à decisão destinado a otimizar as rotas das viaturas de patrulhamento. Sua pesquisa explorou profundamente a aplicação de algoritmos de otimização para criar uma solução de apoio à decisão baseada na divisão de pequenas áreas em quadrantes geográficos. Essa estrutura de divisão é conceitualmente representada por uma matriz com i linhas e j colunas, compreendendo a área de jurisdição G .

O policiamento preditivo, resultante da implementação de algoritmos de otimização, correlaciona a previsão de focos de ocorrências com o tempo, buscando alocar de forma eficiente os agentes de segurança envolvidos nas operações. A previsão de risco baseia-se na análise e predição de séries temporais em cada unidade espacial discretizada. A otimização proposta segmenta a área sob jurisdição da força de segurança em áreas de patrulhamento estrategicamente

orientadas, visando aumentar a frequência da presença policial como meio de repressão à criminalidade.

Leigh, Dunnett e Jackson (2019) , destacam a importância do posicionamento adequado das viaturas em conjunto com a análise em tempo real de dados, não apenas para dissuadir atividades criminosas, mas também para reduzir o tempo de resposta a ocorrências. Nesse sentido, além das técnicas de Estimção da Densidade de Núcleo (*Kernel Density Estimation*, KDE) para a detecção de *hotspots*, Leigh, Dunnett e Jackson (2019) também aplicam a técnica *Tabu Search* baseada em uma função objetivo que combina aleatoriamente pontos, buscando incluir mais *hotspots* nas rotas das viaturas.

No contexto das patrulhas ostensivas, certos locais demandam uma atenção especial e um monitoramento mais rigoroso, tais como estádios, shoppings e escolas. Em algumas situações, a realização de rondas a pé em áreas específicas é necessária, devido a circunstâncias diversas. Ratcliffe et al. (2011) ressaltam que as rondas ostensivas realizadas a pé nessas áreas podem contribuir para a redução de crimes violentos em um nível microespacial. Acredita-se que o esforço intensivo das patrulhas a pé aumenta a certeza da interrupção da atividade criminosa e contribui para a detenção de infratores. No entanto, ainda não existem estudos conclusivos que comprovem a eficácia das rondas a pé e sua correlação com a redução de roubos a transeuntes. Este é um dos aspectos a serem investigados nesta pesquisa.

Capítulo 4

Metodologia

Nesta seção, apresentamos a metodologia adotada neste trabalho, descrevendo as principais etapas e procedimentos que serão seguidos. As etapas do processo de pesquisa incluem:

1. **Análise das Bases de Dados:** Nesta etapa, realizamos uma análise detalhada das bases de dados relevantes para o estudo, avaliando sua qualidade e conteúdo. Isso inclui a identificação de fontes confiáveis de informações sobre ocorrências criminais e dados de patrulhamento.
2. **Preprocessamentos e Integração:** Aqui, descrevemos as ações de preprocessamento realizadas nos dados, incluindo limpeza, transformação e integração das fontes de informação. Além disso, realizamos uma análise exploratória visual de dados para compreender a estrutura inicial das informações coletadas.
3. **Cálculo de Hotspots Baseados em Fundamentos de Análise Topológica:** Descrevemos a metodologia empregada para calcular os hotspots usando princípios de análise topológica, explicando como essa abordagem difere da anterior. Essa abordagem inovadora contribui para uma compreensão mais profunda da distribuição das ocorrências criminais.
4. **Interface de Análise Visual Espaço-Temporal de Crimes e Patrulhamento:** Além das etapas mencionadas acima, desenvolvemos uma interface de análise visual espaço-temporal de crimes e patrulhamento. Essa interface desempenha um papel crucial na tradução dos resultados das análises em informações acionáveis para as autoridades de segurança pública e tomadores de decisão. Detalharemos o desenvolvimento, os recursos e os impactos práticos dessa interface (Capítulo 6)

Cada etapa desempenha um papel crucial no desenvolvimento da pesquisa, contribuindo para a compreensão e identificação de padrões em dados espaciais e temporais de ocorrências criminais.

4.1 Bases de dados

Inicialmente, obtivemos um conjunto de dados fornecidos pela Polícia Militar de Alagoas, provenientes de seu banco de registro de ocorrências e microdados de monitoramento de viaturas. A base de dados de ocorrências apresenta informações detalhadas para cada incidente, incluindo o tipo de crime, localização (cidade, endereço e coordenadas geográficas), data e hora da ocorrência. Destacamos a importância das informações de latitude e longitude, que possibilitam a aplicação de técnicas de discretização espacial com alta precisão.

Dada a heterogeneidade dos dados, tanto em termos do número de crimes quanto em sua distribuição geográfica no Estado de Alagoas, optamos por realizar duas delimitações significativas. Primeiramente, focamos nossa análise nos crimes do tipo roubos a transeuntes, uma vez que diferentes tipos de crimes podem exibir comportamentos distintos no espaço e no tempo. Além disso, considerando as disparidades sociais e econômicas entre os municípios da região, escolhemos concentrar nossa pesquisa na capital do estado, Maceió, a fim de obter uma análise mais representativa e específica.

Em relação à dimensão temporal, baseamos nossas análises nos dados acumulados nos anos de 2019, 2020 e 2021. Essa escolha reflete nossa preocupação em garantir que os dados utilizados sejam mais recentes e representativos da realidade atual.

Os microdados de rastreamento de viaturas, por outro lado, apresentam uma estrutura mais compacta, mas com um grande volume de registros. Este conjunto de dados inclui atributos como identificador do rastreador, data, hora, latitude, longitude e velocidade. No entanto, o rastreamento das viaturas apresentou desafios significativos no processamento devido à alta frequência de registros, uma vez que cada dispositivo gerava uma nova entrada a cada 30 segundos. Esse grande volume de dados, combinado com a cobertura de todo o estado, representou um desafio adicional para a análise, especialmente quando se utiliza algoritmos tradicionais. Além disso, é importante destacar que, somente no município de Maceió, tínhamos mais de 600 dispositivos com registros de monitoramento acumulados no período entre janeiro e outubro de 2021. É relevante mencionar que esses dispositivos incluíam não apenas viaturas, mas também rádios rastreáveis usados por unidades de cavalaria, motocicletas e patrulheiros pedestres.

4.2 Preprocessamentos e integração

Nestas etapas, algumas agregações são realizadas para a obtenção de representações computacionais eficientes que sejam capazes de representar os grandes volumes de dados espaço-temporais das bases de crimes e viaturas. Inicialmente, as localizações das ocorrências criminais são projetadas em suas esquinas mais próximas do grafo de ruas, conforme descrito na Subseção 3.1 e demonstrado na Figura 4.1. Analogamente, este preprocessamento é realizado também na

base de rastreamento de viaturas, visando-se a integração destas duas fontes relevantes de dados no mesmo domínio espacial.

Na sequência do processamento dos dados, é criado para cada esquina uma lista de esquinas vizinhas, cuja distância é parametrizada para uma busca de até 300 metros dentro do grafo de G da cidade, através do peso das arestas que representam essa distância entre cada esquina. Uma vez concluída esta etapa, é aplicada uma função escalar para quantificar a soma de crimes ocorridos em cada vizinhança de cada esquina. Tal processo viabiliza a análise de vizinhanças, o que não seria possível usando as posições fornecidas originalmente pelo par latitude-longitude.

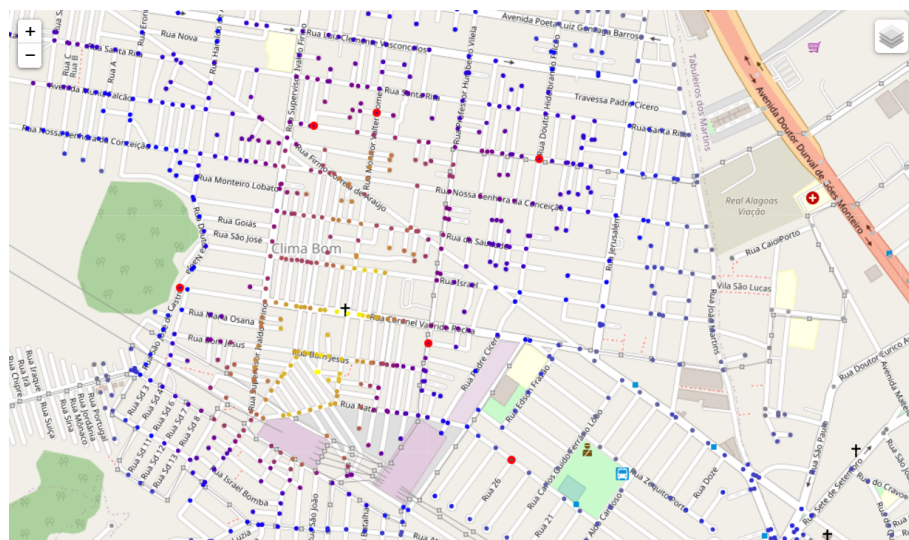


Figura 4.1: Parte de um mapa da cidade de Maceió. Pontos menores em escala de cor ascendendo do azul para o vermelho indicam maior densidade de crimes. Círculos vermelhos indicam os *hotspots*.

O segundo método tem por objetivo viabilizar análise e visualização envolvendo crimes e viaturas a nível de esquina. São geradas duas séries temporais binárias para cada esquina. Cada posição da primeira série representa, binariamente, se houve ou não crime em um certo intervalo de 30 minutos, em um certo dia dentre os 365 dias de um ano, totalizando 17.520 posições por esquina. Apesar de termos adotado uma série com duração de 1 ano, a quantidade de dias poderia ser alterada de acordo com a análise desejada. A segunda série aplica a mesma representação, contudo indicando ou não a passagem de viaturas pela esquina naquele período de tempo. Desta maneira, as representações de crimes e de viaturas deixam de depender da quantidade de ocorrências e de registros de viaturas e passam a ser séries temporais de tamanho fixo, para cada esquina. Intuitivamente, estes vetores permitem identificar, por exemplo, se num certo período de 30 minutos houveram crimes e se houveram passagens de viaturas na esquina correspondente. Estas informações são muito relevantes para analisar correlações espaço-temporais entre ocorrência de crimes e patrulhamento.

4.3 Detecção de hotspots usando heurística de poda nos nós de um grafo

Apresentamos nesta seção uma nova técnica para detecção de hotspots usando heurísticas de poda nos nós do grafo. Após a realização dos preprocessamentos apresentados na Seção 4.2, o conjunto de crimes deve ser dividido em subconjuntos de treino, teste e validação, como é usual em Tarefas de aprendizado de Máquina, porém tendo o tempo como critério de subdivisão, ou seja, o treinamento deve ser realizado em um período anterior T_1 , testado em um período T_2 posterior a T_1 , e validado num período T_3 posterior a T_2 . Neste experimento, os períodos foram separados em trimestres.

Outro pré-processamento pertinente à análise e, necessário à obtenção de resultados relevantes é o processamento espaço-temporal dos crimes numa determinada vizinhança de cada esquina, desta forma, o conjunto de novas covariáveis servirá de apoio para análise da ocorrência de crimes em um agrupamento de vértices próximos geograficamente. Nesta etapa, é realizada a busca dos dados em conjuntos de esquinas que ficam a uma distância máxima no grafo de ruas. O parâmetro padronizado nas análises foi de 300 metros.

Vamos nos basear na definição de que um hotspot é um local onde a ocorrência de crimes é previsível, ou seja, se ocorrer uma certa quantidade de crimes num local em um certo período de tempo, no período de tempo seguinte teremos uma quantidade relativamente similar.

Um componente conexo em um grafo é uma parte na qual todos os vértices estão conectados entre si por caminhos. Após remoções no grafo, diferentes partes podem permanecer ou não conectadas. Identificar os componentes conexos é crucial para compreender a estrutura e a conectividade do grafo resultante. Cada componente é como uma ilha independente, acessível de qualquer vértice para qualquer outro através das arestas restantes. Antes de iniciar o algoritmo de agrupamento topológico de hotspots, é definido um limiar temporal que servirá de parâmetro para realizar a poda em cada validação de períodos. A partir da razão entre o período validado e o período de treino do algoritmo, serão retirados os componentes conexos que não estiverem com valor igual ou superior ao limiar.

Para a execução a separação das regiões de interesse, é delimitado um valor mínimo de crimes numa vizinhança para que ela seja separada como região de preliminar interesse. Executando o algoritmo da Figura 4.2, cada esquina no grafo de ruas, é realizada a poda caso a quantidade de crimes na vizinhança esteja abaixo do do limiar definido. Ao final desta etapa, o grafo está diferente do grafo inicial, delimitando um certo número de regiões no mapa de Maceió/AL. Para refinar as regiões e delimitar melhor transições abruptas de comportamento criminal no tocante à quantidade de crimes, é realizada uma nova varredura no grafo, desta vez através das arestas, verificando se o módulo da diferença de crimes entre as esquinas que compõem as arestas estão superiores ao limiar de arestas, em caso positivo, é realizada a poda da

aresta no grafo.

Uma vez que as etapas são concluídas, restam no grafo G num número n de componentes conexos que, por sua vez serão submetidos à validação nos períodos $t \in \{p_1, p_2\}$ através do algoritmo da Figura 4.3, onde é comparada a evolução da quantidade de crimes, segundo um limiar temporal, para o segundo e terceiro trimestre, quando a condição não é validada, o componente conexo é removido do grafo G , restando ao final um conjunto de componentes conexos do gráfico, que apresentam uma coesão espaço-temporal. Uma organização dos componentes é realizada para se obter em lista, os mais representativos no processo de validação, ou seja, os que mais apresentaram crimes no período p_0 .

Como saída, obtém agrupamentos de esquinas com similaridade espacial, bem como uma coesão no comportamento do crime na região. Os componentes conexos gerados podem indicar as regiões (uma ou mais esquinas) que se adequam às condições do processo de poda. É possível comparar a segmentação feita pelo algoritmo analisando a visualização demonstrada na Figura 4.4, que representa o estado do grafo de ruas antes da poda e avaliar a saída representada na Figura 4.5.

Um resultado da abordagem baseada no agrupamento de esquinas com similaridade

Algoritmo 6: Função Calcular HotSpots

Entrada: Grafo de Ruas (G), função discreta $\mathcal{C}(v, t)$ representando a quantidade de crimes em cada vértice $v \in G$ em um período $t \in \{p_0, p_1, p_2\}$, limiar de crimes na esquina (T_{qc}), limiar de densidade de crimes na Vizinhança (T_{dv}), Limiar de Aresta (T_a)

Saída: Conjunto de Componentes conexos (Hotspots)

CalcularHotspots($G, \mathcal{C}, T_{qc}, T_a, T_t$)

para cada *vértice* v **em** G **faça**

se $\mathcal{C}(v, p_0) < T_{qc}$ **então**

Remover v de G

fim

senão

se $\mathcal{D}(v, t) < T_{dv}$ **então**

Remover v de G

fim

fim

fim

para cada *aresta* e **em** G **faça**

se $|e[0].Crimes(p_0) - e[1].Crimes(p_0)| > T_a$ **então**

Remover e de G

fim

fim

$HotSpots \leftarrow ComponentesConexos(G)$

Saída: $HotSpots$

Figura 4.2: Pseudocódigo da função para segmentação de hotspots através de um grafo de ruas e esquinas.

espacial é diminuir o ruído da análise em modelos sensíveis a dados muito esparsos, como é o caso da análise de crimes. Dentro deste escopo, foram comparadas as séries temporais de dentro dos componentes conexos e foi possível visualizar como cada esquina de um componente se comportava, visualizando a série temporal da esquina e a série temporal do *hotspot*, como pode ser observado na Figura 4.6.

Algoritmo 7: Função Poda HotSpots

Entrada: Grafo de Ruas (G), função discreta $\mathcal{C}(v, t)$ representando a quantidade de crimes em cada vértice $v \in G$ em um período $t \in \{p_0, p_1, p_2\}$, limiar de crimes na esquina (T_{qc}), limiar de densidade de crimes na Vizinhança (T_{dv}), Limiar de Aresta (T_a), Limiar temporal (T_t)

$HotSpots \leftarrow \text{CalcularHotspots}(G, \mathcal{C}, T_{qc}, T_a, T_t)$

Calcule $\tilde{\mathcal{C}}$ para cada componente conexo $c \in HotSpots$, onde $\tilde{\mathcal{C}}(c, t) = \sum_{v \in c} \mathcal{C}(v, t)$

```

para cada  $c$  em  $HotSpots$  faça
    se  $\frac{\tilde{\mathcal{C}}(c, p_1)}{\tilde{\mathcal{C}}(c, p_0)} < T_t$  então
        | Remover  $c$  de  $HotSpots$ ;
    fim
    senão
        | se  $\frac{\tilde{\mathcal{C}}(c, p_2)}{\tilde{\mathcal{C}}(c, p_0)} < T_t$  então
            | Remover  $c$  de  $HotSpots$ ;
        | fim
    fim
fim

```

Ordene HotSpots de acordo com $\tilde{\mathcal{C}}(\cdot, p_0)$.

Saída: HotSpots

Figura 4.3: Pseudocódigo da Poda de um grafo de hotspots para segmentação de componentes conexos com similaridade no comportamento temporal.

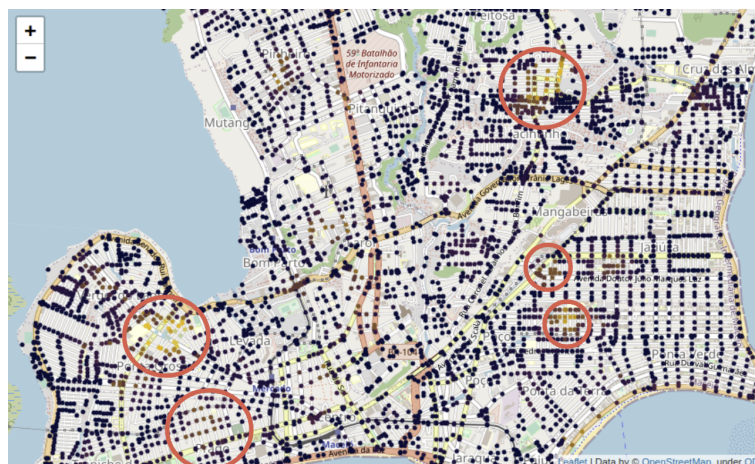


Figura 4.4: Análise de roubos na vizinhança de esquinas de Maceió/AL.

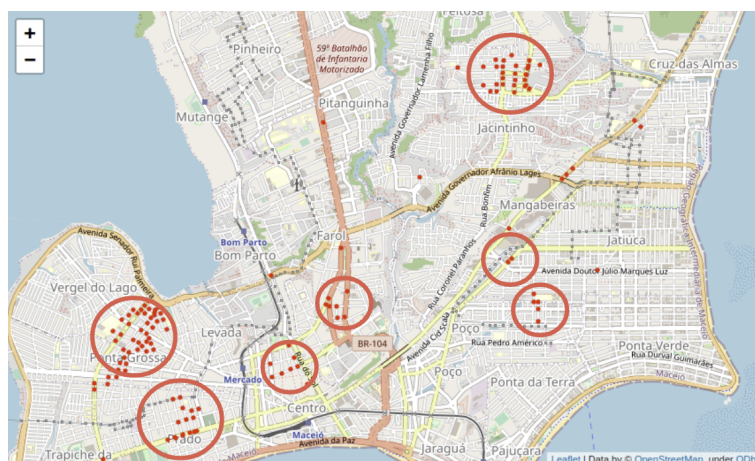


Figura 4.5: Segmentação de hotspots a partir da poda do grafo de Maceió/AL.

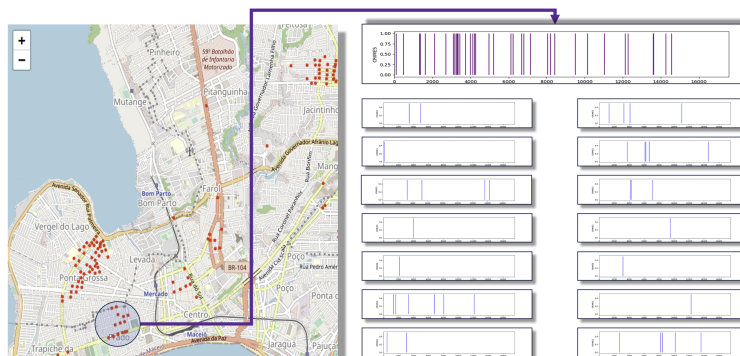


Figura 4.6: Série temporal de um *hotspot*, sendo decomposta em cada esquina.

Capítulo 5

Space-Time Visual Explorer

Neste capítulo apresentamos uma nova ferramenta, denominada *Space-Time Visual Explorer*, capaz de realizar tarefas de análise visual em dados espaço-temporais de segurança pública em Maceió, Alagoas. Em comparação com algumas das ferramentas citadas neste trabalho, o Space-Time Visual Explorer destaca-se pela fusão de fontes de dados de crime e patrulha em uma única plataforma, permitindo a convergência de diferentes medidas e possibilidade de cruzar comportamentos de crime e policiamento no espaço e tempo. Esta ferramenta foi esboçada com o objetivo de realizar as seguintes tarefas analíticas:

- **T1 - Identificação de padrões espaciais de crime.** Localizar espacialmente grupos de esquinas com suas respectivas densidades de crimes em vizinhança.
- **T2 - Filtragem interativa de *clusters* espaciais de hotspots criminais.** Selecionar grupos de esquinas em uma região específica da cidade com alta concentração de crimes, ou alta concentração de crimes na vizinhança.
- **T3 - Filtragem de esquinas por vizinhança, usando ordenação de grafo.** Selecionar intervalos de esquinas vizinhas no grafo de ruas da cidade, utilizando diferentes tipos de indexação e ordenação.
- **T4 - Análise espaço-temporal de crimes.** Visualizar padrões de acúmulos de crime espaciais e temporais num dado ano de análise por esquina.
- **T5 - Identificação de padrões espaciais de patrulhamento.** Identificar geograficamente esquinas e grupos de esquinas com seus respectivos índices de patrulhamento.
- **T6 - Filtragem interativa de *hotspots* de patrulhamento.** Selecionar grupos de esquinas com alto ou baixo índice de patrulhamento.
- **T7 - Análise espaço-temporal de patrulhamento.** Visualizar padrões de patrulhamento espaciais e temporais num dado ano de análise por esquina.

5.1 Componentes visuais

A ferramenta é composta por 4 componentes visuais interativos, exibidos na Figura 5.1, que foram projetados para atender aos requisitos analíticos supracitados, como pode ser demonstrado na Tabela 5.1. A seguir, são detalhadas as funções de cada um dos seguintes componentes visuais: *Index Range Slider*, *Crime and Patrol Map View*, *Index Scatterplot*, *Spatial-Time HeatMap*.

Tabela 5.1: Tarefas Analíticas x Componentes Visuais.

	Tarefas Analíticas						
Componentes Visuais	T1	T2	T3	T4	T5	T6	T7
Index Range Slider	X	X	X	X	X	X	X
Crime and Patrolling Map View	X			X	X		X
Indexed Corners Scatterplot		X	X		X	X	
Spatial-Time Heatmap				X			X

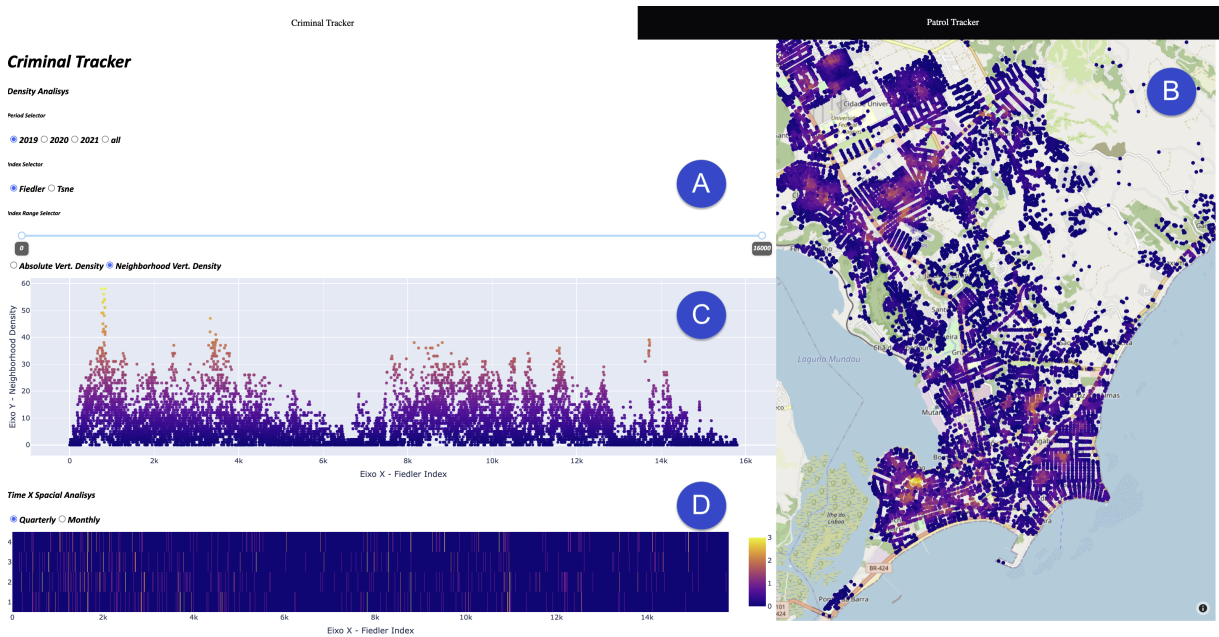


Figura 5.1: Interface do Space-Time Visual Explorer composta por 4 componentes visuais: Index Range Slider (A), Crime and Patrolling Map View (B), Indexed Corners Scatterplot (C), Spatial-Time HeatMap (D)

Através do *Index Range Slider*, representado na Figura 5.1 (A), é possível filtrar grupos de esquinas numa vizinhança de acordo com uma ordenação do grafo de ruas de Maceió. Esta ordenação é calculada usando a ordenação de Fiedler, como apresentado na Seção 2.1.5; ou usando o bem conhecido algoritmo tSNE unidimensional (Maaten e Hinton, 2008). Como estas ordenações tendem a manter nós vizinhos próximos no espaço de índices, espera-se que a seleção

de um intervalo de índices resulte em esquinas próximas geograficamente. Como resultado, os outros três componentes visuais são filtrados em tempo real.

Usando o *Crime and Patrolling Map View* é possível visualizar a distribuição espacial dos dados de ocorrências criminais ou de patrulhamento, diretamente no mapa da cidade, em nível de esquinas. Pontos são plotados sobre a mapa da cidade apenas nas esquinas com ocorrência criminal, ou com passagem de viaturas no período analisado. Um colormap é aplicado nestes pontos, de maneira que a variação na sua escala de cores é definida pelo quantidade de crimes na vizinhança de uma esquina ou pela frequência de passagem de viaturas, viabilizando a localização geográfica de grupos de esquina com maior concentração criminal nas vizinhanças, ou com maior frequência de passagens de viaturas no período analisado. Este componente, porém, não permite selecionar facilmente os hotspots.

Utilizando o *Indexed Corners Scatterplot*, representado na Figura 5.1 (C), é possível visualizar a distribuição de crimes ou de frequência de patrulhamento. No caso de crimes, é possível selecionar dentre duas opções de métricas criminais: a quantidade absoluta de crimes associados a uma esquina; ou quantidade de crimes associados à uma vizinhança da esquina, conforme descrito na subseção 4.2. O eixo horizontal da dispersão está associada ao tipo de índice selecionado na Figura 5.1 (A) que, ao ser alterado também é alterado no *Indexed Corners Scatterplot*. Dessa forma, espera-se que pontos com abscissa parecida representem esquinas próximas no espaço geográfico. O eixo vertical representará a métrica selecionada. Desta forma, os pontos com maior ordenada representam, respectivamente, hotspots ou esquinas muito patrulhadas. Este componente possui recurso interativo de seleção de área, permitindo a seleção de subconjuntos de esquina de interesse, que será filtrada também no *Crime and Patrolling Map View*. Consequentemente, é possível visualizar no mapa qual a localização geográfica das esquinas selecionadas. Este recurso é de suma importância, pois permite a localização tanto de esquinas hotspots isoladas, como de grupos de esquinas que juntas concentram altas concentrações de ocorrências. Na visualização de patrulhamento, é possível observar os picos e vales que representam, respectivamente, grupos de esquina com alta ou baixa volumetria de passagem de viaturas, como ilustra a Figura 6.6.

Utilizando a visualização do *Spatial-Time Heatmap*, demonstrado na Figura 5.1 (D) é possível realizar análise espaço-temporal dos dados de crime e de patrulhamento, em um único componente. Mais especificamente, é possível verificar como os padrões de crimes nas esquinas se comportaram ao longo do espaço e do tempo. Para facilitar a análise, os dados são agregados temporalmente em meses ou trimestres, de acordo com os botões de rádio, e representados no eixo vertical. A forma de indexação no eixo horizontal é análoga a do *Indexed Corners Scatterplot*, sendo determinada pela seleção do índice da esquina na Ordenação de Fiedler ou índice gerado com o tSNE. Padrões sazonais, por exemplo, podem ser identificados com facilidade neste componente.

Capítulo 6

Resultados

6.1 Contextualização

Este capítulo apresenta os resultados obtidos por meio da aplicação do *Space-Time Visual Explorer* em estudos de caso, evidenciando sua eficácia como ferramenta analítica no contexto da segurança pública. Os estudos foram direcionados para a cidade de Maceió, com o intuito de explorar espacial e temporalmente as ocorrências de crimes nos anos de 2019, 2020 e 2021. A abordagem analítica incluiu uma análise exploratória inicial, identificação de padrões temporais, e a execução de tarefas específicas para extração de informações sobre crimes de roubo a transeunte e o patrulhamento policial.

Iniciaremos apresentando a análise exploratória, revelando padrões geográficos significativos das ocorrências criminais na cidade. Em seguida, exploraremos a capacidade do *Space-Time Visual Explorer* em identificar agrupamentos específicos de esquinas relacionadas a comportamentos criminosos, utilizando ferramentas como *Indexed Corner Scatterplot*. Além disso, discutiremos a utilidade do *Index Range Slider* na seleção e análise aprofundada de áreas de interesse, destacando a distribuição de crimes associados às esquinas e suas vizinhanças.

Ao prosseguir, discutiremos o uso do *Spatial-Time Heatmap* para análise temporal trimestral das áreas identificadas, oferecendo uma visão detalhada da evolução dos crimes ao longo do tempo. Posteriormente, exploraremos a distribuição global de passagens de viaturas em Maceió, identificando padrões de patrulhamento em vias principais e secundárias, bem como a aplicação de estratégias interativas, como a remoção de hotspots de patrulhamento, para obter uma visão mais clara das áreas com padrões distintos.

A conclusão do capítulo destacará a importância da integração de índices como o t-SNE e o *Index Range Selector* na análise de microrregiões específicas, apresentando uma análise temporal e espacial detalhada do patrulhamento constante em áreas previamente associadas a altas taxas de criminalidade. Esses resultados fornecem insights valiosos sobre a eficácia do

policiamento constante na mitigação da criminalidade em regiões específicas da cidade. Essa abordagem analítica e os resultados obtidos contribuem significativamente para a compreensão e gestão da segurança urbana.

6.2 Estudo de Caso

Iniciamos a análise exploratória de dados no *Space-Time Visual Explorer* a partir do componente *Crime and Patrol Map View*. Conforme revelado na Figura 6.1, há uma predominância de hotspots no noroeste da cidade (parte alta). Porém, se destacam visualmente também duas manchas relevantes nas regiões mais próximas do litoral (parte baixa, destacadas em círculos vermelhos). Desta maneira, observa-se um padrão geográfico da distribuição dos crimes, conforme almejado na tarefa analítica T1.

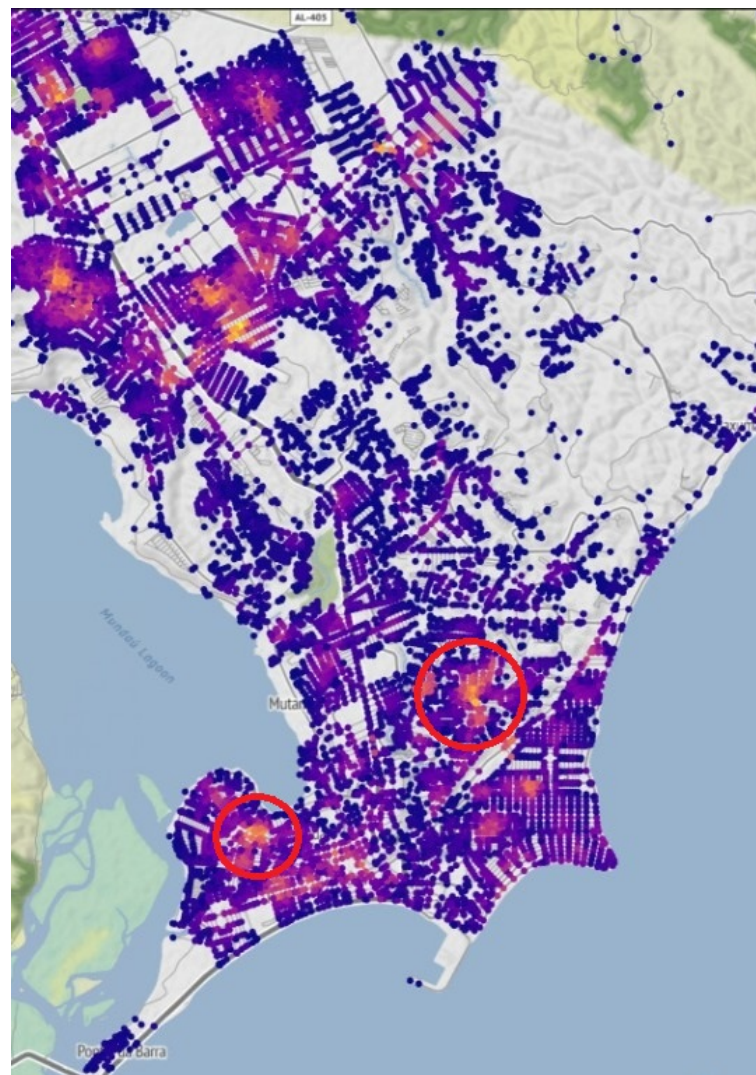


Figura 6.1: Visão da distribuição geográfica dos crimes por vizinhança de cada esquina entre os anos de 2019 e 2021.

É possível selecionar grupos de esquinas através do componente *Indexed Corner Scatterplot (C)*, interativamente, o uso da ferramenta de seleção retangular deste componente permite o filtro de agrupamentos (vizinhanças) que possuem um determinado comportamento que se destaca, seja em quantidade absoluta de crimes por esquina, ou por quantidade de crimes na vizinhança de uma esquina (T1 e T2).



Figura 6.2: Seleção de grupamentos de esquinas a partir da quantidade de crimes ocorridos nas suas vizinhanças.

Identificadas áreas de interesse para análise, faz-se necessário selecioná-las para uma análise mais aprofundada, obtendo-se mais informações e compreensão sobre seus respectivos padrões. Por meio do componente *Index Range Slider (A)*, utilizando o índice de ordenação de Fiedler no grafo (T3), pode ser realizada a separação de esquinas para obtenção da dispersão de crimes associados às esquinas, e crimes associados à vizinhança de uma esquina. De posse da visualização da região selecionada, percebe-se que algumas esquinas específicas concentram mais crimes.

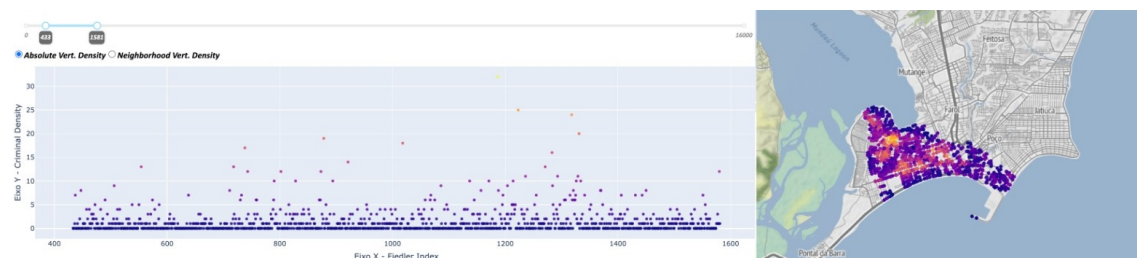


Figura 6.3: Componente visual *Index Scatterplot* da quantidade de crimes numa esquina (Y) pelo respectivo índice de ordenação de Fiedler no grafo (X), apresentado junto com o *Crime and Patrol Map View* com a dimensão geográfica selecionada e colorida de acordo com os crimes de vizinhança das unidades.

Com a visualização anterior, é possível identificar hotspots individuais de esquinas que

se sobressaem, contudo, uma variação da visualização permite agrupar esquinas que individualmente podem não ter um número acentuado de crimes, contudo, ter vizinhanças onde os crimes, de maneira agrupada somam-se e caracterizam alguns *hotspots*. Através do *Index Scatterplot*, é possível selecionar interativamente e aferir a localização espacial destes grupamentos de esquinas (T2), conforme Figura 6.4.

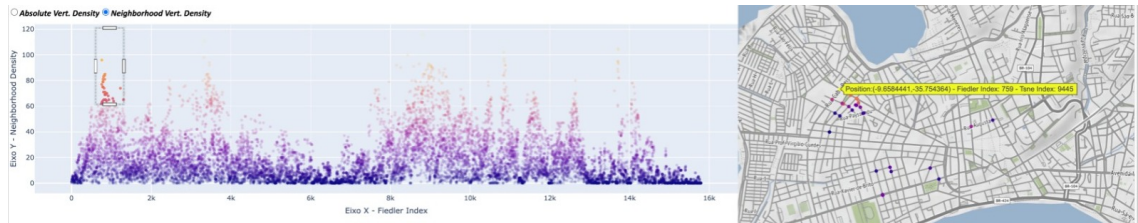


Figura 6.4: *Index Scatterplot* demonstrando distribuições de crimes em vizinhança por esquina, com filtro aplicado no *Crime and Patrol Map View*.

Uma vez compreendido que de fato, naquela localidade agrupavam-se esquinas com certa quantidade de crimes, é importante compreender, ao longo do tempo, como esses crimes ocorreram. Esta análise permite compreender se uma grande quantidade de crimes se concentrou na região num espaço de tempo muito específico, ou se sazonalmente ocorrem crimes na região (T4), entendendo-se essa região como temporalmente consistente. Desta vez, utilizando o índice baseado em t-SNE para obter uma maior precisão na seleção geográfica daquela região no grafo, foram analisados, separadamente, a visualização do componente *Spatial-Time Heatmap* para verificar, trimestralmente, os anos de 2019, 2020 e 2021, conforme demonstrado na Figura 6.5.

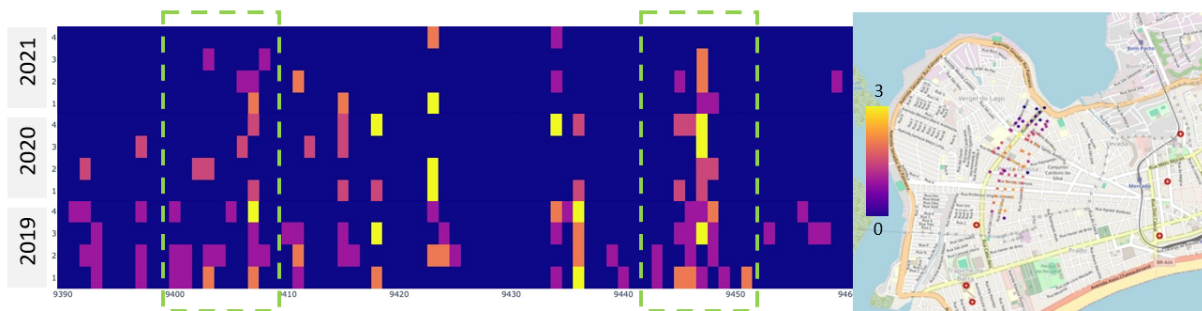


Figura 6.5: Agregações trimestrais por ano do *Spatial-Time Heatmap* para visualização de série de temporal de um hotspot.

Observando dentro do grupo de esquinas, é possível visualizar que duas regiões apresentem uma consistência de ocorrências ao longo do tempo em cada ano. É possível visualizar também que no ano de 2020 o número de casos na região de análise aumentou com relação à 2019, retraindo novamente em 2021, fato este, que gerou um questionamento sobre a redução na quantidade de ocorrências que leva a observar outro elemento, o patrulhamento na região.

Duas análises são necessárias, entender o padrão global de patrulhamento, e entender, na região do *hotspot*, como se deu a distribuição das passagens de viaturas ao longo do tempo.

Sobre patrulhamento, não possuíamos dados de um ano completo. Logo, utilizou-se os dados dos três primeiros trimestres de 2021 para agregação à ferramenta. A região do *hotspot* ficou definida entre os índices t-SNE 9.390 e 9.460, para averiguação de outras métricas referentes ao patrulhamento.

Para medir o patrulhamento, foram coletados e processados dados dos primeiros três trimestres de 2021. Não foram disponibilizados pela PMAL dados de 2020 suficientes para impor uma análise de dados sustentável do período. A métrica utilizada foi um fatiamento do tempo em intervalos de meia hora e foram realizadas verificações indicando se houve passagem de viatura ou não naquela janela de tempo, conforme descrito na subseção 4.2. Ao final, para visualização, foram acumuladas as passagens confirmadas nos intervalos definidos e realizada a agregação desta medida de maneira mensal e trimestral para análise temporal em cada esquina do grafo de Maceió.

Globalmente, verifica-se que esquinas localizadas nas vias principais tornam-se *hotspots* de patrulhamento em razão de sua própria natureza geográfica que predispõe esse comportamento (T5), o que não necessariamente está atrelado ao policiamento em si mas, por exemplo, a passagens de viaturas durante translaços para uma localidade de ocorrência em localidade bem distinta, que enseja a passagem naquela via. Esse comportamento fica bem definido ao analisar o padrão global com o *Crime and Patrol Map View* na Figura 6.6, onde vias arteriais como a BR-104, BR-424 e a BR-316 apresentam muitas passagens de viaturas.

Selecionando um conjunto interativamente através de uma seleção de uma área retangular baixa no componente visual *Indexed Corners Scatterplot (C)*, removeu-se da análise esses hotspots de patrulhamento (T6) e, por consequência, observa-se uma distribuição de patrulhamento mais clara na geografia da cidade, conseguindo visualizar vias secundárias e obter uma visualização que descreve áreas em que na série histórica demonstram um padrão de alto ou baixo patrulhamento de fato, como pode ser demonstrado na Figura 6.7.

Utilizando o índice t-SNE, em conjunto com o *Index Range Selector* para filtrar a microrregião em análise, é possível analisar o comportamento de passagem de viaturas ao longo do tempo e compreender padrões ao longo dos três primeiros trimestres de 2021 disponíveis para análise. Foi possível observar que o agrupamento de esquinas dentro da microrregião que foram responsáveis pela recorrência de crimes na microrregião delimitada teve um patrulhamento acentuado e constante nessas vias durante os três primeiros trimestres do ano de 2021, onde houve uma queda na quantidade de roubos a transeunte no mesmo período (T7). Isto pode ser uma evidência de que o patrulhamento constante mitigou o crime na região. Esse comportamento pode ser visualizado na Figura 6.8.

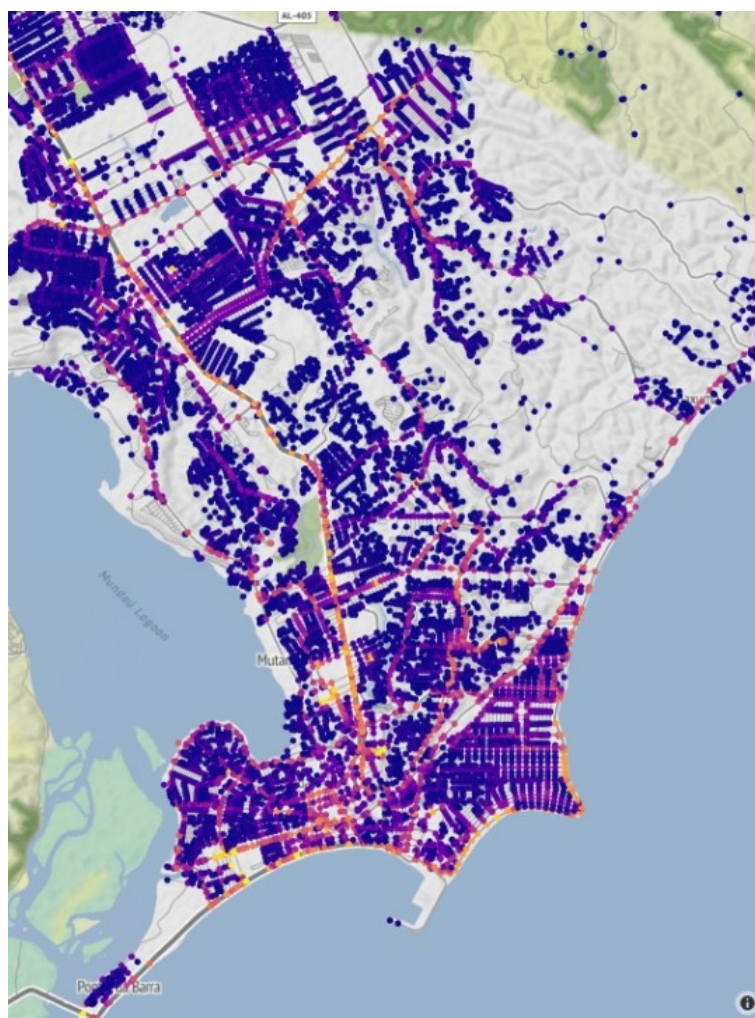


Figura 6.6: Distribuição global de passagens de viaturas em Maceió nos Trimestres de 1 a 3 durante o ano de 2021.



Figura 6.7: Distribuição global de passagens de viaturas em Maceió nos Trimestres de 1 a 3 durante o ano de 2021 após a filtragem que exclui os hotspots.

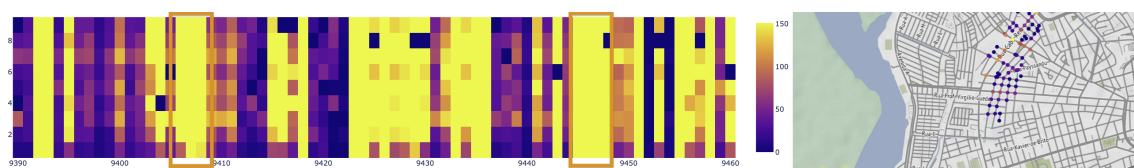


Figura 6.8: Distribuição espacial e temporal do patrulhamento na microrregião durante os três primeiros trimestres de 2021, agregadas por mês.

Capítulo 7

Conclusão e Trabalhos Futuros

Os resultados obtidos no estudo de caso evidenciam que a aplicação do *Crime Tracker* oferece uma abordagem visual e interativa para a identificação de hotspots de crime e patrulhamento, tanto de forma isolada quanto em aglomerados. Esta ferramenta possibilita, de maneira eficaz, a localização de microregiões de interesse, permitindo uma análise aprofundada do comportamento do crime no espaço e no tempo, facilitando comparações com os índices de patrulhamento nas respectivas regiões geográficas ao longo do tempo.

Não obstante, visando potencializar a capacidade analítica e a extração de novos padrões, propõe-se o aprimoramento de algumas funcionalidades específicas:

1. **Ampliação da Captura de Dados Urbanos:** A expansão da variedade de dados urbanos a serem integrados, como informações sobre fluxo de tráfego, iluminação pública e padrões de uso de espaço público, aprimorará a compreensão da dinâmica urbana subjacente aos eventos criminais, possibilitando uma análise mais abrangente e precisa.
2. **Inteligência Artificial para Análise de Imagens:** A implementação de técnicas de inteligência artificial para análise de imagens capturadas, associadas a algoritmos avançados de reconhecimento de padrões, proporcionará uma interpretação automatizada e mais sofisticada das condições visuais nas esquinas. Isso não apenas enriquecerá a análise com informações visuais detalhadas, mas também permitirá a identificação automatizada de padrões relacionados a atividades suspeitas.
3. **Integração de Dados Sociais e Demográficos:** A inclusão de dados sociais e demográficos, como densidade populacional, distribuição etária e perfil socioeconômico, enriquecerá a análise contextual, permitindo uma compreensão mais profunda das correlações entre características demográficas e padrões de criminalidade.
4. **Aprimoramento da Interface de Usuário:** A melhoria da interface do usuário, com ferramentas intuitivas e amigáveis, garantirá uma experiência mais eficiente e acessível.

para os usuários da ferramenta. Isso inclui a implementação de recursos de personalização e visualização mais avançados para facilitar a interpretação e manipulação dos dados.

Essas propostas de desenvolvimento futuro visam não apenas fortalecer a eficácia da ferramenta existente, mas também expandir suas capacidades analíticas, proporcionando uma plataforma mais abrangente e avançada para o entendimento e a gestão de questões relacionadas à segurança urbana.

Referências bibliográficas

- ANDERSON, T. K. Kernel density estimation and k-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, v. 41, n. 3, p. 359–364, 2009. ISSN 0001-4575. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0001457508002340>>.
- BAILEY, T. C.; GATRELL, A. C. et al. *Interactive spatial data analysis*. [S.l.]: Longman Scientific & Technical Essex, 1995. v. 413.
- BERG, M. D.; CHEONG, O. M. van kreveld and mh overmars. *Computational Geometry: Algorithms and Applications*, 2008.
- CAMACHO-COLLADOS, M.; LIBERATORE, F. A decision support system for predictive police patrolling. *Decision Support Systems*, v. 75, p. 25–37, 2015. ISSN 0167-9236. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167923615000834>>.
- CRAGLIA, M.; HAINING, R.; WILES, P. A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies*, v. 37, p. 711–729, 04 2000.
- DAUNI, P. et al. Implementation of haversine formula for school location tracking. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2019. v. 1402, n. 7, p. 077028.
- DOLATSHAH, M.; HADIAN, A.; MINAEI-BIDGOLI, B. Ball*-tree: Efficient spatial indexing for constrained nearest-neighbor search in metric spaces. *arXiv preprint arXiv:1511.00628*, 2015.
- EHLERS, R. S. Análise de séries temporais. *Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná*, v. 1, p. 1–118, 2007.
- GARCIA-ZANABRIA, G. et al. Mirante: A visualization tool for analyzing urban crimes. In: IEEE. *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2020. p. 148–155.
- GARCIA-ZANABRIA, G. et al. Cripav: Street-level crime patterns analysis and visualization. *IEEE Transactions on Visualization and Computer Graphics*, p. 1–1, 2021.
- GARCIA, G. et al. Crimanalyzer: Understanding crime patterns in são paulo. *IEEE Transactions on Visualization and Computer Graphics*, v. 27, n. 4, p. 2313–2328, 2021.
- GRUBESIC, T. H. On the application of fuzzy clustering for crime hot spot detection. *Journal of Quantitative Criminology*, Springer, v. 22, n. 1, p. 77–105, 2006. ISSN 07484518, 15737799. Disponível em: <<http://www.jstor.org/stable/23367481>>.

- HIRSCHFIELD, A.; BOWERS, K. *Mapping and analysing crime data: Lessons from research and practice*. [S.l.]: CRC Press, 2001.
- KOSHY, T. *Discrete mathematics with applications*. [S.l.]: Elsevier, 2004.
- LEIGH, J.; DUNNETT, S.; JACKSON, L. Predictive police patrolling to target hotspots and cover response demand. *Annals of Operations Research*, Springer, v. 283, n. 1, p. 395–410, 2019.
- LODHA, S. K.; VERMA, A. K. Spatio-temporal visualization of urban crimes on a gis grid. In: *Proceedings of the 8th ACM international symposium on Advances in geographic information systems*. [S.l.: s.n.], 2000. p. 174–179.
- MAATEN, L. Van der; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. 11, 2008.
- NETO JOSÉ FLORÊNCIO E SANTOS, E. M. d. e. V. C. A. de Q. Mskde - usando marching squares para criar rapidamente mapas de hotspots de crimes de alta qualidade. In: *2016 29ª Conferência SIBGRAPI sobre Gráficos, Padrões e Imagens (SIBGRAPI)*. [S.l.: s.n.].
- NIE, K. et al. A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: A case study from wuhan, china. *Sustainability*, v. 2015, p. 2662–2677, 03 2015.
- RATCLIFFE, J. Crime mapping: spatial and temporal challenges. *Handbook of quantitative criminology*, Springer, p. 5–24, 2010.
- RATCLIFFE, J. H.; MCCULLAGH, M. J. Hotbeds of crime and the search for spatial accuracy. *Journal of geographical systems*, Springer, v. 1, n. 4, p. 385–398, 1999.
- RATCLIFFE, J. H. et al. The philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology*, Wiley Online Library, v. 49, n. 3, p. 795–831, 2011.
- ROSEN, K. H.; KRITHIVASAN, K. *Discrete mathematics and its applications: with combinatorics and graph theory*. [S.l.]: Tata McGraw-Hill Education, 2012.
- TONRY, M. H.; GHIROTTI, J. C.; MORRIS, N. *Policimento moderno*. [S.l.]: EdUSP, 2003.
- VARGAS, M. A.; ALVES, N. G.; MREJEN, M. Ciência, tecnologia e inovação em tempos de pandemia: implicações da covid-19. *Cadernos do desenvolvimento*, v. 16, n. 28, p. 145–172, 2021.