

CrimAnalyzer: Compreendendo os padrões de crime em São Paulo

Germain Garcia, Jaqueline Silveira, Jorge Poco Membro, IEEE, Afonso Paiva, Marcelo Batista Nery, Claudio T. Silva
Fellow, IEEE, Sergio Adorno, Luis Gustavo Nonato Membro, IEEE

Resumo — São Paulo é a maior cidade da América do Sul, com índices de criminalidade que refletem seu tamanho. O número e o tipo de crimes variam consideravelmente na cidade, assumindo padrões distintos dependendo das características urbanas e sociais de cada localidade.

Trabalhos anteriores se concentraram principalmente na análise de crimes com a intenção de descobrir padrões associados a fatores sociais, sazonalidade e atividades urbanas rotineiras. Portanto, esses estudos e ferramentas são mais globais, no sentido de que não são projetados para investigar regiões específicas da cidade, como bairros, avenidas ou áreas públicas. Ferramentas capazes de explorar locais específicos da cidade são essenciais para que especialistas do domínio realizem suas análises de forma ascendente, revelando como características urbanas relacionadas à mobilidade, comportamento de pedestres e presença de infraestruturas públicas (por exemplo, terminais de transporte público e escolas) podem influenciar a quantidade e o tipo de crimes. Neste artigo, apresentamos o CrimAnalyzer, uma ferramenta de análise visual que permite aos usuários estudar o comportamento de crimes em regiões específicas de uma cidade. O sistema permite que os usuários identifiquem pontos críticos locais e o padrão de crimes associados a eles, ao mesmo tempo em que mostra como os pontos críticos e os padrões de crime correspondentes mudam ao longo do tempo. O CrimAnalyzer foi desenvolvido a partir das necessidades de uma equipe de especialistas em criminologia e lida com três grandes desafios: i) flexibilidade para explorar regiões locais e compreender seus padrões de criminalidade; ii) identificação de pontos críticos de criminalidade espacial que podem não ser os mais prevalentes em termos de número de crimes, mas que são importantes o suficiente para serem investigados; e iii) compreensão da dinâmica dos padrões de criminalidade ao longo do tempo. A eficácia e a utilidade do sistema proposto são demonstradas por comparações qualitativas e quantitativas, bem como por estudos de caso conduzidos por especialistas na área, envolvendo dados reais. Os experimentos demonstram a capacidade do CrimAnalyzer na identificação de fenômenos relacionados à criminalidade.

Termos de índice — Dados criminais, dados espaço-temporais, análise visual, fatoração de matriz não negativa

1 INTRODUÇÃO

DESDE meados da década de 1970, a sociedade brasileira vivenciou uma processo de transição da ditadura militar para a democracia.

Com essa transição política, esperava-se que os conflitos fossem cada vez mais solucionados, reduzindo a prevalência da violência. Isso não ocorreu. De fato, a transição foi acompanhada por uma explosão de conflitos, muitos dos quais associados à criminalidade urbana. Ainda não há consenso entre os cientistas sociais sobre as razões que explicam essas tendências na evolução da criminalidade e da violência na sociedade brasileira, em particular nas grandes cidades [1].

Entre as explicações que surgem com mais frequência está o esgotamento dos modelos tradicionais de política de segurança. Em relação a este último aspecto, é inegável que a criminalidade não só cresceu, como também se tornou mais violenta e modernizada. Em contrapartida, os órgãos responsáveis pela lei e pela ordem (por exemplo, a polícia e o sistema de justiça criminal) não acompanharam essas tendências. A lacuna entre a dinâmica do crime e da violência e a capacidade do Estado de contê-los dentro do Estado de Direito aumentou. Portanto, a introdução de instrumentos modernos de gestão da ordem pública e de contenção da criminalidade é imperativa para tornar as políticas de segurança pública mais eficientes, não apenas

somente em países de ¹ mas também em qualquer grande cidade em desenvolvimento São Paulo.

O Mapeamento Criminal, um ramo dos Sistemas de Informação Geográfica (SIG) dedicado a explicar o comportamento espaço-temporal dos crimes, surgiu como um campo de pesquisa para apoiar criminologistas em seu processo analítico, alavancando a importância da geografia local como determinante dos tipos de crimes e da taxa de ocorrência em uma determinada região [48]. A capacidade de identificar e visualizar pontos críticos de crimes e a capacidade de filtrar atributos relacionados a crimes para revelar informações específicas, como roubos em áreas comerciais ou a sazonalidade de roubos de automóveis em determinados bairros, está entre os principais componentes de uma abordagem de mapeamento criminal [38].

A maioria das ferramentas existentes desenvolvidas para mapeamento da criminalidade concentra-se na detecção de hotspots, ou seja, áreas com alto número de incidentes criminais [14]. Embora mecanismos sofisticados tenham sido propostos para detectar hotspots [15], a busca por uma alta prevalência de crimes acaba negligenciando locais onde certos tipos de crimes são frequentes, mas não suficientemente intensos para serem considerados estatisticamente significativos [51]. Além disso, a maioria das técnicas permite apenas mecanismos rudimentares para analisar um componente importante das atividades ilícitas: a evolução temporal dos crimes e seus padrões. De fato, os recursos de visualização para análise temporal disponíveis na maioria dos sistemas de mapeamento da criminalidade são muito restritivos, impedindo os usuários de realizar consultas elaboradas e exploração de dados [3].

Há ainda outro aspecto importante a ser considerado no contexto do mapeamento da criminalidade: as especificidades das áreas urbanas em análise. São Paulo, por exemplo, apresenta um dos maiores índices de criminalidade.

• Germain Garcia, Jaqueline Silveira e Afonso Paiva estão no ICMC-São Carlos, Brasil. E-mail: {germaingarcia, alva.jaque}@usp.br, USP, apneto@icmc.usp.br •

Marcelo Batista Nery trabalha no CEPID-FAPESP e no Instituto de Estudos Avançados – Programa Cidades Globais. E-mail: mbnery@gmail.com • Sergio

Adorno trabalha no NEV-USP, São Paulo, Brasil. E-mail: marsadorno@usp.br • Jorge Poco está na Fundação

Getúlio Vargas, Brasil e Universidade Católica San Pablo. E-mail: jorge.poco@fgv.br • Cláudio Silva está na New York University, EUA.

E-mail: csilva@nyu.edu • Luis Gustavo Nonato trabalha no ICMC-USP, São Carlos, Brasil e na New York University, EUA. E-mail: gnonato@icmc.usp.br

1. São Paulo é um estado e uma cidade. Neste artigo, qualquer momento que não especificar explicitamente, São Paulo se referirá à cidade.

taxas no mundo, pelo menos uma ordem de magnitude maior do que cidades como Nova York e São Francisco, tornando soluções de mapeamento de crimes baseadas em glifos, como LexisNexis, completamente inadequadas para analisar CrimeMapping⁴ crimes em São Francisco. em São Paulo. No entanto, o padrão de crimes muda consideravelmente em São Paulo, mesmo entre regiões geograficamente próximas, exigindo soluções analíticas personalizadas para revelar os pontos críticos locais e os padrões de criminalidade correspondentes. Essas soluções locais também devem ser capazes de revelar a dinâmica dos pontos críticos ao longo do tempo. Esses recursos não estão disponíveis atualmente na maioria das ferramentas de mapeamento criminal.

Este trabalho apresenta o CrimAnalyzer, uma nova ferramenta de análise visual personalizada para auxiliar na análise de atividades criminosas em áreas urbanas com as características de São Paulo, ou seja, altas taxas de criminalidade com grande variabilidade no padrão de crimes, mesmo em regiões geograficamente próximas. O CrimAnalyzer possibilita uma série de visualizações interligadas, personalizadas para revelar padrões de crimes e sua evolução ao longo do tempo, auxiliando especialistas na área em seu processo de tomada de decisão e fornecendo diretrizes não apenas para ações repressivas, mas sobretudo preventivas, fortalecendo o planejamento e a implementação de ações institucionais, especialmente da polícia.

Em colaboração com uma equipe de especialistas na área, projetamos funcionalidades de análise visual que permitem aos usuários selecionar e analisar regiões de interesse em termos de seus pontos críticos, padrões de criminalidade e dinâmica temporal. Além disso, o sistema proposto oferece recursos para que os usuários explorem mais profundamente locais específicos para compreender seus crimes prevalentes e seu comportamento ao longo do tempo. Além disso, o CrimAnalyzer implementa uma metodologia baseada em Fatoração de Matriz Não Negativa [27] para identificar pontos críticos com base não apenas no número de crimes, mas também na taxa de ocorrência.

Em resumo, as principais contribuições deste trabalho são: • Uma nova metodologia para identificar focos de criminalidade com base não apenas no número de crimes, mas também em sua variação e taxa de recorrência.

- Um mecanismo de análise visual que permite aos usuários realizar consultas espaciais e temporais visualmente para entender padrões e dinâmicas temporais de crimes.
- CrimAnalyzer, uma ferramenta assistida por visualização que integra o maquinário analítico em um conjunto de visualizações vinculadas. O CrimAnalyzer opera em regiões espaciais alvo para descobrir informações relevantes da região como um todo e também de seus locais individuais.
- Um conjunto de estudos de caso que revelam fenômenos interessantes sobre a dinâmica do crime em São Paulo, apoiando hipóteses e teorias levantadas por especialistas no assunto e descritas na literatura.

2 TRABALHOS RELACIONADOS

A literatura sobre análise de crimes é extensa, abrangendo desde estatística e ciência de dados até visualização e SIG. Em termos gerais, os métodos de análise de crimes podem ser agrupados em duas categorias principais: abordagens georreferenciadas e não georreferenciadas. As últimas abordagens, não georreferenciadas, baseiam-se em mecanismos matemáticos e computacionais, como mineração de dados [12], [55], otimização [51], aprendizagem de máquina [52], [54], estatística [35] e visualização de dados [16], [53], para identificar padrões de crime, comportamento criminoso e também a consistência da justiça criminal.

A seguir, porém, nos concentramos em técnicas georreferenciadas desenvolvidas para mapeamento de crimes que estão mais intimamente relacionadas à nossa abordagem. Para melhor contextualizar nossa metodologia, dividimos as técnicas georreferenciadas em dois grupos: técnicas centradas em hotspots e técnicas de identificação de padrões criminais espaço-temporais. Deve ficar claro que há uma sobreposição considerável entre esses grupos, o que significa que uma técnica centrada em hotspots também pode se basear em padrões espaço-temporais para potencializar sua análise, mas o foco principal dessa técnica é, de fato, a identificação de hotspots.

Centrado em hotspots. Identificar hotspots de crimes é uma tarefa importante no contexto do mapeamento do crime [8], [14], [15]. Embora alguns trabalhos se baseiem em Krigagem [33], a abordagem mais comum para identificação de hotspots é uma combinação de Estatísticas de Varredura Espacial [25] e Estimativa de Densidade de Kernel (KDE) [10], usando nuvens de pontos, mapa de densidade ou abordagens baseadas em grade como recursos de visualização [22], [32], [45]. Conforme apontado por Hart e Zandbergen [21], definir corretamente os parâmetros de um KDE não é fácil e uma escolha imprecisa de parâmetros pode levar a resultados errôneos ou imprecisos que superestimam ou desconsideram os locais de hotspots [9]. Outro problema com técnicas baseadas em KDE é que locais que apresentam atividades criminosas regulares, mas não intensas, dificilmente são apontados como hotspots. Para evitar os problemas acima, nossa abordagem se baseia na Fatoração de Matrizes Não Negativas para detectar hotspots, evitando assim o ajuste de parâmetros e, ao mesmo tempo, sendo capaz de “capturar” locais difíceis de serem identificados como hotspots por métodos baseados em KDE.

Identificação de padrões espaço-temporais. Além da identificação de pontos críticos de criminalidade, a análise e a visualização de padrões temporais e espaciais de criminalidade também são de grande importância no mapeamento de crimes [7], [13], [29], [32], [42]. Ratcliffe [41] e Townsley [47], por exemplo, incorporam a análise aorística [39], [40] em seus sistemas de visualização de pontos críticos para identificar padrões espaço-temporais importantes de crimes. A análise aorística leva em consideração a incerteza do momento exato em que um evento ocorreu ao examinar a incidência geral de crimes ao longo do tempo. Lukasczyk et al. [30] fornecem uma perspectiva topológica da evolução temporal de pontos críticos com base em um gráfico de Reeb espaço-temporal construído a partir de um mapeamento KDE. Embora interessantes, as técnicas descritas acima ainda são incipientes em revelar claramente os padrões espaço-temporais de criminalidade e suas dinâmicas. Nossa abordagem, em contraste, combina uma série de recursos visuais intuitivos a partir dos quais se pode identificar claramente os padrões de criminalidade e sua evolução temporal em locais específicos.

Existem diversas técnicas espaço-temporais que se baseiam em métodos de agrupamento para agrupar eventos criminais espacial e/ou temporalmente semelhantes, a fim de identificar padrões. Esses métodos podem ser organizados em duas categorias: aqueles que se baseiam em algoritmos de agrupamento convencionais e aqueles que se baseiam em Mapas Auto-Organizáveis (SOM) para identificar padrões. Métodos baseados em agrupamento extraem vetores de características de atributos criminais espaciais e temporais e agrupam esses atributos por meio de k-means [2], [46] ou agrupamento por vizinho mais próximo [23], [28].

O principal objetivo das técnicas que se baseiam em SOM é a identificação de similaridades entre atributos de crimes. Chen et al. [11], em colaboração com o Departamento de Polícia de Tucson, propuseram um sistema de visualização espaço-temporal denominado COPLINK, que combina árvores hiperbólicas, SIG e SOM em uma ferramenta analítica unificada. Andrienko et al. [4] utilizam uma matriz SOM para alavancar uma estrutura analítica visual e explorar similaridades espaço-temporais entre eventos. Hagenauer et al. [20] ampliaram as abordagens anteriores para explorar a evolução espaço-temporal dos padrões, além de seus aspectos demográficos e socioeconômicos.

2. communitycrimemap.com
3. maps.nyc.gov/crime/
4. crimemapping.com

atributos. Para compreender padrões entre os tipos de crime, o SOM também serviu de base para o sistema de análise espaço-temporal de crimes proposto por Guo e Wu [19], que se baseia em uma infraestrutura de visualização chamada VIS-STAMP [18], que integra redução de dimensionalidade e coordenadas paralelas na análise de padrões de crime. O SOM apresenta problemas bem conhecidos, como a configuração adequada de pesos, número de nós e sobreajuste [26].

Além disso, as técnicas baseadas em SOM descritas acima não integram a detecção de hotspots como parte do sistema, deixando de lado um importante componente de análise no contexto do mapeamento do crime.

3 DESAFIOS, CONJUNTO DE DADOS E ANALÍTICA TAREFAS

Durante dezoito meses, interagimos com dois especialistas em ciências sociais cuja pesquisa se concentra em análise criminal. Um dos sociólogos é um pesquisador renomado no estudo da violência na América do Sul. O outro sociólogo é especialista em segurança pública e ciências sociais aplicadas a questões urbanas, com formação em SIG e vasta experiência em análise espaço-temporal da criminalidade. Em parceria com o Departamento de Polícia do Estado de São Paulo, a equipe de especialistas construiu um conjunto de dados (detalhado na Seção 3.2) contendo sete anos de registros criminais em São Paulo. Eles nos contataram para desenvolver uma ferramenta de análise visual para auxiliar na compreensão e análise dos dados.

Nomenclatura. Antes de detalharmos mais o problema, os requisitos levantados a partir da interação com os especialistas da área e o conjunto de dados, vamos primeiro definir algumas nomenclaturas que serão empregadas ao longo do manuscrito.

– Sítio é a menor unidade territorial dada na discretização espacial. Em nosso contexto, os sítios são definidos como as unidades censitárias de São Paulo, cada uma contendo de 250 a 350 residências e/ou estabelecimentos comerciais.



– Região é um conjunto de terrenos, que pode corresponder a um bairro inteiro, a uma parte específica de um bairro ou até mesmo a um grupo de terrenos adjacentes a uma rua ou avenida. O encarte à direita mostra um exemplo de uma região e seus terrenos correspondentes.

– Pontos críticos são conjuntos de locais dentro de uma região com atividade criminosa relevante. O significado exato de "relevante" ficará claro quando apresentarmos o mecanismo que projetamos para a detecção de pontos críticos. Os locais avermelhados na imagem inserida correspondem aos pontos críticos na região em questão.

– O tipo de crime refere-se ao tipo de atividade criminosa, que vai de roubo a lesão corporal (morte, crimes sexuais e crimes relacionados a drogas não estão incluídos em nosso estudo).

– O padrão de criminalidade representa a prevalência de um grupo de tipos de crime em uma determinada região ou local. Em outras palavras, se dissermos que o padrão de criminalidade em um conjunto de locais é roubo, furto de carro e ataque a estabelecimento comercial, queremos dizer que os três tipos de crime são os mais prevalentes nesses locais.

3.1 Análise de Problemas

Realizamos várias rodadas de reuniões e entrevistas com os especialistas para identificar os principais desafios envolvidos na análise de dados criminais. Após diversas interações, chegamos às seguintes questões:

• Analisando as características e a dinâmica dos crimes em regiões

específicas da cidade. A partir de sua experiência e interação com policiais do departamento de polícia, os especialistas conjecturam que o tipo e a dinâmica dos crimes vêm mudando ao longo dos anos, principalmente em regiões específicas da cidade.

Além disso, o tipo de crime pode mudar mesmo em regiões próximas umas das outras, dependendo das características urbanas de cada região. A principal dificuldade para realizar essa análise sem uma ferramenta de análise visual é consultar adequadamente o conjunto de dados, o que pode ser um trabalho demorado e exaustivo. Muitas vezes, um grande número de imagens é gerado como resultado, e o trabalho de analisá-las se torna impossível.

Além disso, crimes de alta prevalência ofuscam a presença de crimes menos frequentes, que também podem ser de interesse, exigindo ferramentas específicas para permitir uma análise adequada. Dadas as dificuldades, os especialistas têm realizado suas análises com foco em apenas um ou dois tipos de crime, considerando a cidade como um todo ou analisando grandes áreas que servem como unidades administrativas dentro da cidade. Essa análise ampla dificulta a validação (ou negação) de hipóteses e conjecturas de natureza local. • **Identificar pontos críticos de criminalidade em uma determinada região.** A identificação de pontos

críticos de criminalidade está entre as tarefas mais importantes na análise de crimes e sua dinâmica. Os pontos críticos são geralmente identificados como locais que possuem um número maior que a média de registros criminais [14]. No entanto, locais criminosos que não são tão prevalentes em termos de número de eventos criminais, mas apresentam atividades criminosas que merecem uma análise especial, tendem a não ser detectados quando uma abordagem "frequentista" é empregada para identificar pontos críticos de criminalidade. Devido à falta de mecanismos mais sofisticados, o número de registros criminais tem sido o principal mecanismo empregado pelos especialistas na identificação de pontos críticos. Por isso, foi necessário propor um novo método para detecção de pontos críticos que atendessem às restrições descritas. Esse requisito foi enfatizado pelos especialistas da área.

• **Compreendendo e comparando padrões de criminalidade.** Especialistas acreditam que locais e hotspots dentro da mesma região podem apresentar diferentes padrões de criminalidade. Uma questão nesse contexto é saber se o padrão de criminalidade varia de um local (ou hotspot) para outro na mesma região. Em caso afirmativo, os especialistas gostariam de entender como os tipos de crime se distribuem e como evoluem ao longo do tempo. Os especialistas buscavam uma solução que lhes permitisse fazer essas comparações de forma intuitiva.

Os desafios acima apontam para uma solução analítica visual dotada de funcionalidades que permitam selecionar facilmente regiões de interesse, ao mesmo tempo em que disponibiliza recursos para auxiliar na análise da localização, padrões e evolução temporal do crime. Seguimos um processo de design que envolveu especialistas na maioria das etapas do desenvolvimento [31], redesenhando procedimentos, componentes e funcionalidades de acordo com o feedback e as demandas dos especialistas.

~

3.2 Dados de Roubo, Arrombamento e Furto em São Paulo O

conjunto de dados compilado pelos especialistas consiste em registros criminais fornecidos pela polícia de São Paulo. Foram fornecidos apenas atos criminosos de roubo, arrombamento e furto, excluindo homicídio, homicídio, crime relacionado a drogas e agressão sexual.

Cada registro contém o número de identificação da unidade censitária (local) onde ocorreu o crime, o tipo de crime,



Os tipos de crime abrangem 121 categorias, e os 10% mais frequentes correspondem a cerca de 80% do total de crimes. O encarte à direita mostra o histograma dos 10% mais frequentes, identificando os três mais frequentes: roubo a pedestres, furto de veículos e furto a pedestres. Para facilitar a análise, os especialistas dividiram os dados originais em três categorias independentes: roubo de veículos (inclui carros, motocicletas, caminhões, etc.) com 295.081 ocorrências, furto em geral, com 587.885 ocorrências, e uma terceira categoria com todos os outros tipos de roubo e arrombamento, com 691.954 ocorrências.

Embora o número de tipos de crimes seja bastante grande, os crimes pelos quais os especialistas do domínio estão interessados não são tão numerosos, variando de 3 a 5. Outros tipos de crimes são escassos o suficiente para serem analisados individualmente e não requerem uma ferramenta de visualização sofisticada para interpretá-los. Além disso, alguns crimes podem ser agrupados em categorias, uma alternativa sugerida pelo especialista e incorporada ao CrimAnalyzer. Em outras palavras, em cada um dos três subconjuntos de dados, os especialistas classificaram e agruparam os tipos de crimes de acordo com sua importância.

3.3 Tarefas Analíticas

Após identificar os principais desafios enfrentados pelos especialistas e compreender como os dados foram estruturados, conduzimos uma nova série de entrevistas para levantar questões a serem investigadas. Ficou claro que os especialistas estão interessados em compreender a dinâmica da criminalidade na cidade, analisando a variação dos padrões de criminalidade ao longo do espaço e do tempo. A partir dos processos iterativos com os especialistas, compilamos a seguinte lista de tarefas analíticas:

- **Seleções interativas (T1):** Como as regiões espaciais de interesse

podem ser selecionadas interativamente? É possível tornar a seleção interativa de regiões flexível o suficiente para selecionar desde pontos únicos até bairros inteiros e avenidas específicas?

- **Padrões de criminalidade na cidade (T2):** Quais são os padrões de criminalidade em regiões e locais específicos? Como os padrões de criminalidade mudam do centro para áreas residenciais e periferias? E quanto aos padrões ao longo das principais avenidas, ruas e rodovias?

- **Dinâmica dos crimes ao longo do tempo (T3):** Como os tipos de crimes evoluíram, ao longo do tempo, em regiões específicas da cidade? Mais especificamente, os padrões de criminalidade mudaram em regiões específicas ao longo dos anos? Os tipos de crimes são sazonais?

- **Padrões de criminalidade e pontos críticos no espaço (T4):** Quais são os pontos críticos em uma determinada região? Quais são seus padrões de criminalidade? Quão diferentes (se é que a diferença existe) são os padrões de criminalidade em diferentes pontos críticos dentro da mesma região?

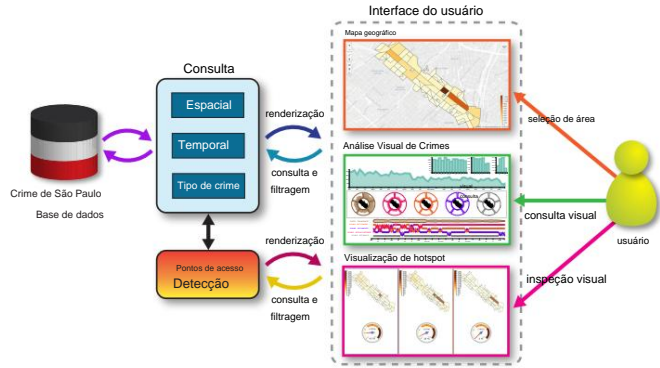


Fig. 1. Visão geral do pipeline do sistema CrimAnalyzer.

- **Padrões e pontos críticos de criminalidade ao longo do tempo (T5):** Os pontos críticos de criminalidade mudaram ao longo do tempo em uma determinada região? Os padrões de criminalidade mudaram ao longo do tempo em um determinado ponto crítico?

Como mencionado anteriormente, a falta de mecanismos interativos para selecionar regiões de interesse, combinados com técnicas gerais de análise e visualização, impediu que especialistas no domínio explorassem livremente os dados para verificar hipóteses e conjecturas. O primeiro passo para viabilizar recursos analíticos mais poderosos é o desenvolvimento de uma ferramenta de seleção interativa adequada, sendo este o objetivo do T1.

Também ficou claro durante as entrevistas que é importante ir além de resumos de alto nível e chegar a análises individuais de locais e pontos críticos. Analisar dados em diferentes escalas também é essencial para entender como os padrões variam ao longo do espaço e do tempo. Por exemplo, o padrão de crimes e pontos críticos pode mudar ao longo dos meses e em diferentes dias da semana. Esse fato está relacionado ao T3 e requer agregação e filtragem específicas de dados para ser devidamente abordado.

As tarefas analíticas T2 e T3 estão relacionadas ao problema de compreensão dos diferentes padrões de criminalidade na cidade e sua evolução ao longo do tempo, conforme discutido na Seção 3.1, enquanto as tarefas T4 e T5 estão associadas ao problema de análise de pontos críticos, também discutido na Seção 3.1. Para serem adequadamente abordadas, essas tarefas exigem mecanismos específicos para detectar pontos críticos e também recursos visuais para explorá-los e compreendê-los.

Entre nossos objetivos está a integração de métodos de seleção interativos e ferramentas dedicadas de análise visual, permitindo que especialistas em domínio realizem análises confirmatórias e exploratórias. Além disso, alguns especialistas em domínio não possuem formação em ciência da computação, portanto, o sistema deve ser o mais simples e intuitivo possível. No entanto, simplicidade e expressividade devem ser equilibradas para tornar o sistema capaz de suportar análises espaço-temporais em diferentes escalas, além de revelar pontos críticos e padrões de criminalidade não triviais.

3.4 O Sistema CrimAnalyzer Com base nos

requisitos e tarefas analíticas descritos na Seção 3.3, desenvolvemos o CrimAnalyzer, um sistema para explorar dados espaço-temporais sobre crimes em locais específicos. O CrimAnalyzer permite recursos visuais simples, porém atraentes, para consultar, filtrar e visualizar dados sobre crimes. Os recursos visuais são suportados por um maquinário matemático e computacional desenvolvido para extrair e refinar informações, de modo a apresentá-las visualmente de forma intuitiva e significativa. Os módulos e a arquitetura do sistema são ilustrados na Figura 1. Os usuários consultam visualmente o conjunto de dados interagindo com um mapa e selecionando uma região de interesse, bem como interagindo com as diferentes visualizações vinculadas que compõem o sistema.

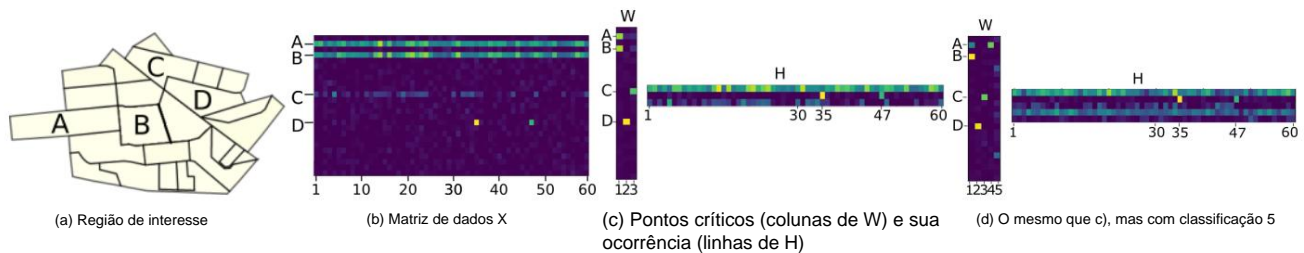


Fig. 2. a) Região de interesse. b) Matriz de dados contendo informações sobre crimes das regiões em a). As linhas correspondem aos locais, enquanto as colunas são fatias de tempo. Quanto mais escura a cor, mais próximo de zero o número de crimes. c) Decomposição de NMF de nível 3 a partir de X. d) NMF de nível 5 a partir de X.

4 MODELO DE IDENTIFICAÇÃO DE HOTSPOT

Conforme discutido na Seção 3, a identificação de hotspots é uma das tarefas mais importantes para a análise de crimes. Aqui, os hotspots têm uma conotação mais geral do que em trabalhos anteriores, correspondendo a locais onde a atividade criminosa é alta, mas também a locais onde o número de crimes não é grande, mas frequente o suficiente para merecer uma análise detalhada. Por exemplo, em uma determinada região, locais cujo número de crimes é muito maior do que em quaisquer outros locais são claramente hotspots importantes. No entanto, a região também pode conter um local específico onde os crimes são frequentes, mas ocorrem em magnitude muito menor se comparados aos mais proeminentes. A região também pode conter locais onde os crimes não são frequentes, mas apresentam picos em períodos específicos. Consideramos os três fenômenos diferentes como hotspots, buscando identificar locais onde os crimes são frequentes e em grande número, locais onde os crimes são frequentes, mas não em grande número, e locais onde os crimes não são frequentes, mas ocorrem em grande número em períodos específicos. Os diferentes comportamentos criminosos serão discutidos mais adiante.

ilustrado na Seção 4.2.

Análise de locais individuais. Existem muitas alternativas para identificar pontos críticos, conforme discutido na Seção 2. Embora variem em termos de complexidade, as técnicas existentes normalmente se baseiam na comparação de informações estatísticas para identificar pontos críticos. Os pontos críticos podem ser identificados como pontos específicos ou como unidades de área, dependendo de como os dados são organizados, delegando à visualização a tarefa de revelar adequadamente os pontos críticos. O problema com essa abordagem é que crimes que ocorrem em pequena magnitude ou em intervalos de tempo isolados tendem a não ser estatisticamente significativos, dificilmente sendo apontados como pontos críticos.

Outro problema é que vários locais podem ser identificados como hotspots, mas sua relação temporal permanece obscura. Por exemplo, os locais podem ser correlacionados temporalmente, o que significa que crimes são cometidos nesses locais nos mesmos intervalos de tempo. Faz sentido agrupar locais correlacionados temporalmente em um único hotspot, mas calcular hotspots individualmente e agrupá-los de acordo com correspondências temporais não é fácil e envolve o uso de limites para decidir quais locais devem ser agrupados.

Análise de grupos de locais. Em vez de analisar locais individualmente, pode-se recorrer a técnicas que identificam diretamente grupos de locais como hotspots. Uma alternativa simples é extrair características dos locais e aplicar um esquema de agrupamento para agrupar locais semelhantes em hotspots (ver Seção 2). No entanto, o problema de extrair características significativas que caracterizem locais espacial e temporalmente é bastante complexo, principalmente devido à escassez dos dados de criminalidade. No decorrer do nosso desenvolvimento, tentamos diferentes alternativas para definir vetores de características de criminalidade espaço-temporais, variando de janelas de tempo cumulativas simples a métodos mais sofisticados baseados em coeficientes wavelet de gráfico [49], mas não conseguimos obter resultados que atendessem aos nossos requisitos.

Para contornar as dificuldades apontadas acima, optamos por uma abordagem baseada na Fatoração de Matrizes Não Negativas (NMF) [27], que funcionou muito bem para nós na identificação de hotspots de acordo com nossas necessidades.

4.1 Fatoração de Matrizes Não Negativas Antes de

apresentar os detalhes de como adaptamos a NMF para operar em nosso trabalho, vamos revisar brevemente os principais conceitos e ideias envolvidos em uma análise NMF. Uma matriz $m \times n \times X$ é considerada não negativa se todas as entradas em X forem maiores ou iguais a zero ($X \geq 0$).

O objetivo da NMF é decompor X como um produto $W \cdot H$, onde W e H são matrizes não negativas com dimensões $m \times k$ e $k \times n$, respectivamente (os papéis de m , n e k ficarão claros na Seção 4.2). Em termos matemáticos, o problema pode ser formulado da seguinte forma:

$$\underset{W, H}{\operatorname{argmin}} \|X - WH\|_F^2 \quad \text{sujeito a } W, H \geq 0 \quad (1)$$

Uma solução para o problema de minimização (Equação 1) fornece um conjunto de vetores base w_i , correspondendo às colunas de W , e um conjunto de coeficientes h_j , correspondendo às colunas de H , de modo que cada coluna x_j de X seja escrita como a combinação linear $x_j = \sum_i h_{ji} w_i$, (ou $x_j = W h_j$). Em outras palavras, para cada coluna em X , temos uma coluna correspondente em H , cujas entradas são coeficientes associados às colunas (vetores base) de W . A representação matricial abaixo (Equação 2) ilustra a relação entre as colunas de X e H , bem como os coeficientes em H e os vetores base.

vetor em W .

$$\begin{matrix} \bar{y} & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} \\ \bar{y} & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} \\ \bar{y} & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} \end{matrix} = \begin{matrix} \bar{y} & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} \\ \bar{y} & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} \\ \bar{y} & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} & \dots & | & \bar{y} \end{matrix} \begin{matrix} w_1 & w_2 & \dots & w_k \\ h_1 & h_2 & \dots & h_n \\ h_k \end{matrix} \quad (2)$$

Existem dois aspectos importantes em uma decomposição NMF que serão amplamente explorados no contexto de detecção de hotspot, a saber, aproximação de baixa classificação e esparsidade. A aproximação de baixa classificação é responsável pelo fato de que a matriz base W geralmente tem uma classificação muito menor do que a matriz original X , o que significa que (as colunas de) X são representadas usando apenas alguns vetores base. Conforme detalhado na próxima subseção, contamos com a aproximação de baixa classificação para definir o número de hotspots, ou seja, ao definir a classificação de W , também definimos o número de hotspots. A esparsidade significa que as matrizes base e coeficiente contêm muitas entradas iguais (ou próximas) a zero, o que naturalmente força que apenas informações relevantes de X sejam mantidas em W e H . Esse fato é importante para identificar locais específicos dentro de um hotspot e as fatias de tempo em que cada hotspot aparece.

4.2 Identificando Hotspots com NMF

Contamos com NMF para identificar hotspots, sua taxa de ocorrência e "intensidade". A matriz X a ser decomposta como o produto

W -H abrange informações sobre crimes em uma região de interesse específica. Especificamente, cada linha em X corresponde a um local da região e cada coluna a uma fatia de tempo. Para facilitar a discussão, apresentamos a abordagem proposta usando um exemplo sintético. A Figura 2(a) mostra uma região composta por 25 locais, e geramos dados sintéticos de crimes em 60 fatias de tempo, representando meses ao longo de cinco anos.

Para os locais denotados como A e B, extraímos 60 amostras de uma distribuição normal com média 8 e variância 4, garantindo que A e B sejam correlacionados, ou seja, quando o número de crimes em A é grande, o mesmo acontece com B (o número de crimes em B é gerado pela perturbação dos valores de A usando uma distribuição aleatória uniforme com valores entre -3 e 3). Essa construção simula duas regiões com alta prevalência de crimes ao longo do tempo. Os crimes no local denotado como C na Fig. 2(a) seguem uma distribuição normal com média 1 e variância 4, correspondendo a um local onde os crimes não são em grande número, mas acontecem com bastante frequência. Finalmente, para o local D, extraímos 60 amostras de uma distribuição normal com média 0 e variância 0,25, exceto para as fatias de tempo 35 e 47, onde definimos o número de crimes igual a 15 e 10, respectivamente, simulando um local onde os crimes não são frequentes, mas acontecem em grande número em fatias de tempo específicas. Para todos os outros locais, associamos 60 amostras retiradas de uma distribuição normal com média 0 e variância 0,25. Os valores para todos os locais são arredondados para o número inteiro mais próximo e os valores negativos são zerados. A Figura 2(b) ilustra a matriz X construída a partir dos dados sintéticos descritos acima. Observe que a dinâmica simulada do crime é claramente visível em X.

Dada uma matriz $m \times n \times X \geq 0$, uma decomposição NMF de X resulta em matrizes $W \geq 0$ e $H \geq 0$. Na prática, o posto de W é significativamente menor que m e n, ou seja, $k = \text{posto}(W) \ll m, n$. Aqui, as colunas de W correspondem aos pontos de acesso, enquanto as entradas nas linhas de H indicam a "intensidade" do ponto de acesso em cada intervalo de tempo. A Fig. 2(c) ilustra as matrizes W e H obtidas da matriz X na Fig. 2(b) usando uma decomposição NMF com posto $k = 3$. Observe que as entradas na primeira coluna (mais à esquerda) de W têm valores próximos de zero em quase todos os lugares, exceto nas entradas correspondentes aos locais A e B. Portanto, o hotspot derivado da primeira coluna de W destaca os locais A e B como os relevantes. A alta prevalência de crimes nessas regiões pode ser vista claramente na primeira linha (superior) da matriz H, que tem a maioria de suas entradas com valores diferentes de zero. A segunda coluna de W é praticamente nula, exceto na entrada correspondente ao local D, onde os crimes não são frequentes, mas acontecem com alta intensidade em fatias de tempo específicas. Observe que a segunda linha de H tem basicamente duas entradas diferentes de zero, correspondendo exatamente às fatias de tempo 35 e 47, quando o local D enfrenta um grande número de crimes. Finalmente, a última coluna de W dá origem a um hotspot que destaca o local C, onde os crimes são frequentes, mas em menor magnitude quando comparados a A e B.

A incidência e a intensidade dos crimes em C são vistas claramente na terceira linha (inferior) de H.

Pode-se argumentar que os resultados apresentados na Fig. 2(c) funcionaram tão bem porque, sabiamente, definimos a classificação de W como $k = 3$ e que, na prática, é difícil encontrar um valor adequado para a classificação. Para responder a essa pergunta, a Fig. 2(d) mostra o resultado da fatoração da matriz X, definindo a classificação de W como $k = 5$. Observe que a principal diferença entre as fatorações com classificação $k = 3$ e classificação $k = 5$ é que a primeira coluna de W na Fig. 2(c) foi dividida em duas colunas na fatoração com classificação $k = 5$, dando origem às colunas 1 e 4 de W na Fig. 2(d). No entanto, a primeira coluna ainda indica a correlação entre A e B, que portanto não é completamente perdida devido à presença da coluna 4. A coluna mais à direita de W na Fig. 2(d) é principalmente ruído e representa sites com pouca atividade criminosa, o que é

atestado pela linha inferior de H na Fig. 2(d), que é quase nula. Portanto, aumentar k tende a dividir pontos críticos significativos, ao mesmo tempo em que cria alguns pontos ruidosos e não tão importantes, que podem ser facilmente identificados a partir de quase zero linhas em H.

Melhorando a identificação de hotspots. A maioria das entradas na matriz H está próxima de zero, mas não é zero, exigindo um limite para decidir se um hotspot ocorre ou não em uma determinada fatia de tempo. Usar limites é sempre inconveniente, principalmente para usuários inexperientes. Para evitar o uso de limites, binarizamos a matriz H usando o algoritmo de Otsu [36], considerando que um hotspot aparece em uma determinada fatia de tempo se sua entrada correspondente na versão binarizada da matriz H for 1.

O exemplo sintético discutido acima mostra que os hotspots gerados a partir da NMF atendem aos requisitos do nosso problema, justificando nossa escolha da NMF como modelo matemático para lidar com o problema. Dentre as diferentes versões da NMF, optamos pela fatoração de matriz esparsa não negativa proposta por Kim e Park [24], que permite impor a esparsidade em W e H simultaneamente.

Concluimos esta seção dizendo que, até onde sabemos, esta é a primeira vez que o NMF é usado como um mecanismo para detectar pontos críticos no mapeamento de crimes.

Comparação com estatísticas espaciais O Getis-Ord G_i* [17], [34] é um método de detecção de hotspot bem conhecido disponível na caixa de ferramentas Indicador Local de Associação Espacial (LISA) [5]. G_i* opera medindo a variação da autocorrelação espacial local i relata um valor p e um sobre uma região de interesse. G_i* escore z para cada local na região de interesse, marcando como pontos críticos aqueles com escores z grandes estatisticamente significativos (valores p baixos).

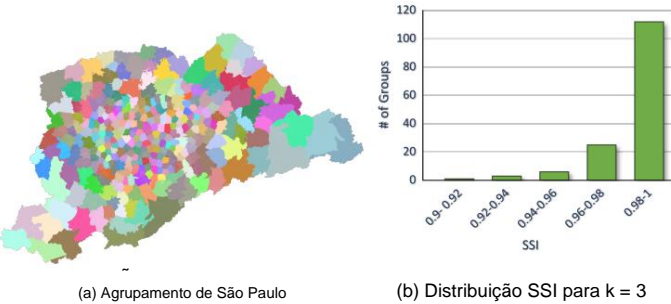


Fig. 3. (a) Divisão de São Paulo em 300 grupos e (b) Distribuição de SSI nessas regiões. NMF e G detectam os mesmos hotspots na maioria das regiões. casos.

Para realizar uma comparação quantitativa entre NMF, agrupamos as unidades censitárias em 300 regiões, conforme mostrado na Figura 3(a). As regiões foram calculadas aplicando k-means às coordenadas do centroide das unidades censitárias. Como há

Considerando que existem cerca de 30.815 unidades censitárias, definir o número de clusters como 300 tende a gerar grupos com cerca de 100 unidades nas áreas mais densas da cidade. O índice de Sokal-Sneath (SSI), uma medida de similaridade de classificação de dados binários bem conhecida [44], é empregado para comparar os hotspots resultantes do NMF com $k = 3$ (o valor de classificação padrão em nosso sistema) com os obtidos por G com um nível de confiança de 99% (confiamos no G da biblioteca PySAL Python [43]). A implementação disponível em

Especificamente, atribuímos cada local a uma das quatro categorias (rótulos):

- P: se o local for um hotspot para NMF e G
- F: se o local for um hotspot detectado por NMF, mas não por G
- N: (correspondência positiva);
- S: (correspondência negativa)

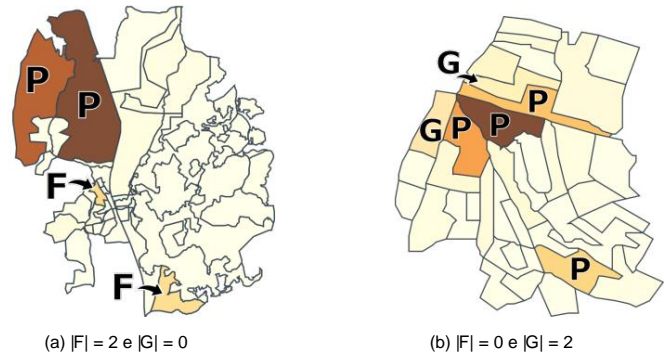


Fig. 4. Comparação qualitativa entre NMF e G. (a) Região onde detecta mais NMF detecta mais pontos de acesso G. (b) região onde há pontos de acesso G.

- G: se o site for um hotspot detectado por G • N: se eu, mas não pela NMF; o site não for um hotspot para ambos os métodos (correspondência negativa).

A medida de similaridade SSI é então calculada como:

$$SSI = \frac{2|P|+2|N|}{2|P|+|F|+|G|+2|N|},$$

onde | | denota a cardinalidade. Um SSI = 1 significa que os pontos críticos detectados por ambos os métodos em uma determinada região correspondem exatamente.

O histograma mostrado na Fig. 3(b) reúne valores de SSI de todas as 300 regiões. Observe que os valores de SSI são maiores que 0,90, a maioria deles situando-se na faixa [0,98, 1,00], demonstrando a boa correspondência entre NMF e G. De fato, a maioria dos locais indicados também são capturados por NMF. No entanto, em cerca de 200 regiões, o NMF detectou alguns pontos críticos a mais do que o G.

A Fig. 4 ilustra situações típicas em que NMF e G encontraram ambos os pontos críticos em alguns locais. Na Fig. 4(a), os hotspots NMF e G são rotulados como P (os rótulos na Fig. 4 estão de acordo com a classificação usada pelo SSI: quanto mais escuro o local, mais crimes ele apresenta), mas o NMF capturou dois hotspots extras, rotulados como F (unidades não rotuladas pertencem à categoria N). Observe que o código de cores indica que os locais F são de fato regiões com prevalência de criminalidade, embora não tenham sido capturados pelo G.

Na Fig. 4(b), em contraste, detecta dois hotspots a mais que o NMF (sites G). Observe que o código de cores das unidades marcadas como G na Fig. 4(b) indica que os crimes nessas regiões não são tão intensos quanto nos hotspots P. O que aponta os sites G como hotspots é que o número de crimes nessas regiões faz com que G sejam vizinhos de unidades onde o número de crimes é alto (hotspots "reais"), então a integração do kernel empregada por G acaba sendo contaminada pelos sites vizinhos onde os crimes são prevalentes. Em outras palavras, os sites G são apontados como hotspots devido à sua proximidade com os sites P. Os sites apontados como F na Fig. 4(a) não foram capturados por G porque estão isolados no meio de unidades sem crimes. Portanto, além de não exigir uma discretização de grade, o NMF tende a capturar hotspots de forma mais consistente, sendo uma alternativa atraente às abordagens estatísticas convencionais.

O valor k (classificação NMF) impacta a medida do SSI. Realizamos as comparações variando k = 3, ..., 10, obtendo um SSI médio maior que 0,98 para k = 3, 4, 5, mas ligeiramente melhor para k = 3. Esse resultado nos motivou a definir k = 3 como o valor padrão em CrimAnalyzer.

| Faceta | Filtro | Tarefa |
|---------------------------------------|-------------------------------------|--------|
| Tipo de espaço-tempo | Tipo de espaço-tempo T1 T2 T3 T4 T5 | |
| Visualização do mapa | | |
| Visualização de pontos de acesso | | |
| Visão Temporal Cumulativa | | |
| Visão Temporal Global | | |
| Visualização do tipo de classificação | | |
| Visão do tipo radial | | |

TABELA 1

Propriedades de visualização e suas tarefas analíticas (Seção 3.3).

5 DESIGN VISUAL

Esta seção descreve os componentes visuais do CrimAnalyzer.

A Fig. 5 ilustra o sistema baseado na web, que compreende um Menu de Controle (a), seis visualizações interativas (bg) e um widget de filtro (h). A Tabela 1 mostra as propriedades de cada visualização. Por exemplo, a Visualização de Mapa mostra a faceta espacial e a Visualização de Tipo de Classificação, as facetas temporal e de tipo de crime. As colunas sob a categoria de filtro mostram como interagir com cada visualização.

Algumas visualizações permitem restringir o espaço, o tempo e os tipos de crime. O design dos recursos visuais foi orientado pelas tarefas analíticas descritas na Seção 3.3. Na Tabela 1, as colunas sob a categoria de tarefa indicam a relação de cada visualização com as tarefas analíticas.

Por exemplo, a visualização do tipo de classificação e a visualização do tipo radial são responsáveis por todas as tarefas analíticas.

Na área de análise de crimes, visualizações sempre foram utilizadas para exibir dados; no entanto, melhorias ou novos designs em relação às visualizações existentes são necessários. Por exemplo, a Visualização por Tipo de Classificação é uma visualização alternativa inovadora nesse contexto, que se mostra eficaz para elucidar a dinâmica de diferentes tipos de crimes ao longo do tempo em locais específicos da cidade. Embora essa metáfora visual seja bem conhecida pela comunidade de visualização, ela nunca foi usada para análise de crimes.

A seguir, descrevemos cada componente visual, começando pelo Menu de Controle (veja Fig. 5).

5.1 Menu de controle

O menu de controle possui três opções: conjunto de dados, discretização temporal (ou seja, meses ou dias) e número de hotspots (classificação da decomposição NMF). Conforme mostrado na Figura 5(a), utilizamos o conjunto de dados "Roubo" com discretização mensal e três hotspots (k = 3) na maior parte de nossas análises. A decomposição NMF é realizada utilizando a biblioteca Python Nimfa [56], que avalia e escolhe automaticamente o valor de classificação k. No entanto, esse processo automatizado é computacionalmente custoso, limitando seu uso em um aplicativo de análise visual interativa como o CrimAnalyzer. Portanto, optamos por permitir que o usuário defina k manualmente.

5.2 Visualização do

Mapa: Este é o primeiro componente utilizado pelos usuários para iniciar o processo analítico, onde o usuário pode definir a região de interesse. Esta visualização é composta por um mapa geográfico e um mapa coroplético para codificar o número de crimes cometidos em cada local da região.

Além disso, os usuários podem aplicar zoom e panorâmica ao mapa.

Seleção de região: Os usuários podem definir uma região de interesse 1) clicando no mapa (para selecionar um local), 2) desenhando uma polilinha (para selecionar avenidas ou ruas, por exemplo), 3) desenhando um polígono (para selecionar um bairro inteiro) ou 4) informando o endereço de um local. Os desenhos podem ser expandidos para incluir outros locais no bairro. Por fim, o CrimAnalyzer define a região calculando os locais que cruzam o desenho. Na Figura 5(b), podemos ver como uma região é representada.

Seleção do local: Durante a exploração, quando uma região já foi definida, esta visualização pode ser usada para filtragem espacial (por exemplo, para

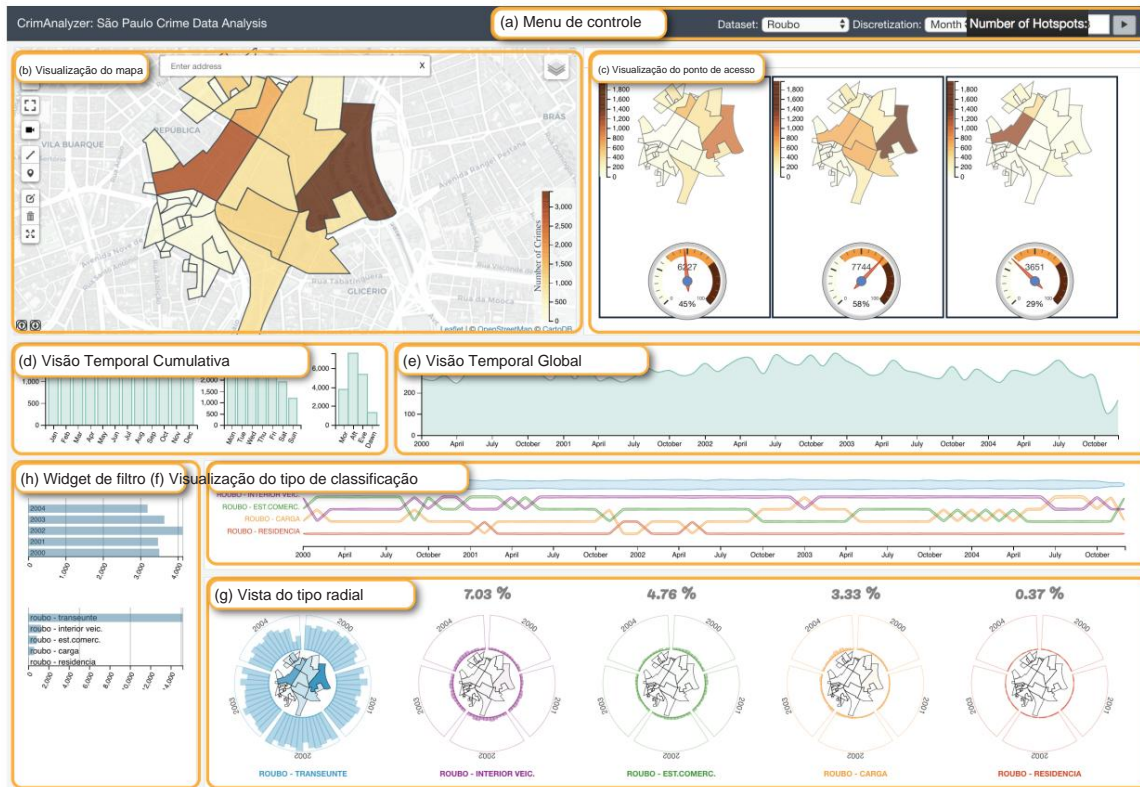


Fig. 5. Sistema CrimAnalyzer: as visualizações interativas espaciais e temporais permitem a exploração de regiões locais enquanto revelam seus padrões criminais ao longo do tempo.

foco em um site específico). Esta operação é realizada clicando em um site, que é destacado pelo mapeamento de uma textura para o local correspondente área.

Filtragem: Quando outras visualizações fazem filtragem espacial (ou seja, selecionam um site), o site correspondente é destacado alterando sua textura. Quando um filtro de tempo ou tipo é ativado por outras visualizações, nosso mapa coroplético é recalculado usando os dados filtrados.

5.3 Visualização de

Pontos de Acesso Um componente importante da nossa abordagem é a identificação de pontos de acesso. Na Seção 4.2, explicamos como a Fatoração de Matriz Não Negativa foi usada para revelar pontos de acesso. Nesta visualização, usamos vários mapas para representar a distribuição espacial de cada ponto de acesso. Os usuários podem especificar o número de pontos de acesso no Menu de Controle. Abaixo de cada ponto de acesso (ver Fig. 5(c)), há um widget de medidor que descreve o número de crimes no ponto de acesso (o número superior no medidor), a taxa temporal de ocorrência do ponto de acesso (a porcentagem inferior no medidor) e a relevância desse ponto de acesso em todo o conjunto de crimes (o ponteiro do medidor). A importância do hotspot é calculada por uma função $f: [0, 1] \times [0, 1] \rightarrow [0, 1]$ que atribui um valor a cada par (taxa_de_crimes, frequência_de_crimes), onde taxa_de_crimes denota o número de crimes no hotspot dividido pelo total de crimes e frequência_de_crimes é o número temporal de ocorrências do hotspot (calculado para a matriz binarizada H) dividido pelo número total de pontos de acesso. Na verdade, f é simplesmente uma interpolação bilinear no quadrado unitário onde $f(0, 0) = 0$, $f(0, 1) = 0.5$, $f(1, 0) = 0.7$, $f(1, 1) = 1$. Com essa distribuição de valores, damos mais relevância aos pontos críticos onde o número de crimes é maior.

Seleção: Uma seleção de pontos de acesso filtra os crimes por espaço e tipo.

Todas as outras visualizações são recalculadas para corresponder ao ponto de acesso selecionado.

Filtragem: Filtrar os crimes usando outras visualizações (por exemplo, espaço, tempo ou tipo) não afeta esta visualização. Se quisermos recalculer os pontos críticos com base nos dados filtrados, por exemplo, um tipo específico de crime, precisamos clicar no botão "Pontos críticos" após realizar a filtragem de dados.

5.4 Visão Temporal Global Esta visão

fornece uma visão geral do número de crimes cometidos durante todo o período de tempo, com base em um gráfico de linhas com uma área preenchida entre o valor dos dados e a linha de base zero (consulte a Fig. 5(e)).

Seleção de tempo: Nesta visualização, podemos restringir a análise a um intervalo de tempo específico, que pode ser definido deslizando o dedo sobre um retângulo na Visualização Temporal Global. Apenas períodos de tempo contínuos podem ser selecionados. A próxima visualização nos permitirá selecionar múltiplos intervalos de tempo. Todas as visualizações (exceto o ponto de acesso que precisa ser recalculado) são afetadas e ajustadas automaticamente de acordo com a seleção de tempo.

5.5 Visão temporal cumulativa Esta visão

usa um gráfico de barras para apresentar o número de crimes acumulados por mês, dia e período do dia (ver Fig. 5(d)).

Nesta visualização, podemos observar alguns padrões de intervalos de tempo não contínuos. Isso também é muito útil para comparar fins de semana ou dias úteis.

Filtragem: Quando outras visualizações são usadas para filtrar o conjunto de dados, os dados filtrados também são exibidos na Visualização Temporal Cumulativa global, permitindo assim uma análise comparativa (ver Fig. 9).

5.6 Visão de Tipo de

Classificação Esta visão descreve três informações relevantes em uma única metáfora: evolução do tipo de crime, classificação do tipo de crime e número de crimes em cada período. Conforme mostrado na Figura 5(f), cada tipo de crime

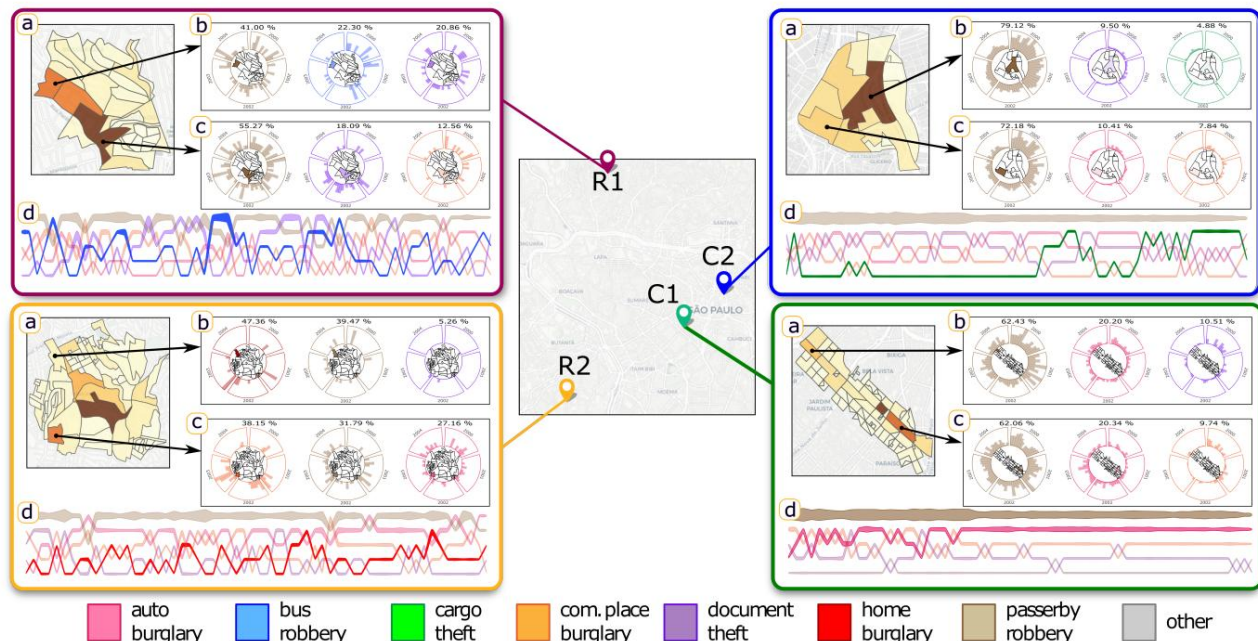


Fig. 6. Resumo das atividades criminosas e padrões correspondentes em quatro regiões diferentes de São Paulo. Os padrões de criminalidade podem mudar substancialmente entre as regiões e também ao longo do tempo.

é representado por uma polilinha. A posição vertical, em cada passo de tempo, codifica a relevância em relação aos demais. Além disso, a largura da linha é proporcional ao número de crimes pertencentes a ela.

Filtragem: quando um filtro é ativado em outras visualizações, a visualização de classificação é recomputada usando os dados filtrados.

5.7 Exibição do tipo radial

Nesta exibição, estamos usando vários gráficos de barras com um layout radial. Cada gráfico representa um tipo de crime diferente. Por exemplo, na Figura 5(g), temos cinco tipos de crime. Além disso, o número no topo de cada gráfico mostra a porcentagem de cada tipo de crime. Cada gráfico é dividido em setores, onde cada setor é composto por 12 barras representando os meses de cada ano.

Seleção do tipo de crime: Clicar em um gráfico filtra os dados para um tipo de crime específico. Dessa forma, os usuários podem concentrar sua análise nos tipos de crime mais prevalentes. Os tipos de crime selecionados são representados por uma linha tracejada.

Seleção de tempo: fornecemos recursos de interatividade em cada gráfico para permitir a comparação entre o mesmo mês em anos diferentes e o mesmo mês em diferentes tipos de crimes.

Filtragem: quando o conjunto de dados é filtrado, cada gráfico é recomputado para representar os dados filtrados.

5.8 Widget de Filtro Este

widget é composto por um histograma de tempo e tipo de crime. Por exemplo, a Figura 5(h) resume nossos dados em cinco anos (2000-2004) e cinco tipos de crime. Além disso, usamos esses histogramas para filtrar nossos dados. Clicando em uma barra, podemos remover um ano ou um tipo de crime. Essa filtragem afeta toda a interface. Embora a maioria dos recursos visuais apresentados não sejam inéditos individualmente, muitos deles (como a visualização por hotspot, a visualização por ranking e a visualização por radial) não são triviais no contexto do mapeamento da criminalidade. Mais importante ainda, a combinação de todos eles permite múltiplas análises simultaneamente, revelando padrões de criminalidade interessantes, como mostrado na próxima seção.

6 ESTUDOS DE CASO

Esta seção apresenta três estudos de caso que demonstram a eficácia do CrimAnalyzer na abordagem das tarefas analíticas apresentadas na Seção 3.3. O primeiro estudo de caso aborda as tarefas analíticas T1, T2 e T3, enquanto o segundo se concentra na análise de hotspots e está relacionado a T4 e T5. O terceiro estudo de caso visa estabelecer um paralelo entre a atividade criminosa em São Paulo e alguns fenômenos relacionados à criminalidade relatados na literatura (relacionados a T3). Em todos os estudos de caso, exceto quando explicitamente indicado, utilizamos o bloco de dados de roubo e furto, conforme descrito na Seção 3.2, com discretização mensal.

6.1 Comparando os padrões de criminalidade na cidade (T1, T2, T3)

O objetivo deste estudo de caso é analisar os padrões de criminalidade em diferentes regiões da cidade, a fim de compreender como eles mudam de acordo com as características urbanas. Além disso, também investigamos a evolução temporal dos padrões de criminalidade em diferentes regiões.

Para realizar o estudo, selecionamos quatro áreas em São Paulo: duas no centro da cidade, identificadas como C1 e C2 na Figura 6, e duas em áreas residenciais, identificadas como R1 e R2 na Figura 6. C1 é um distrito financeiro, que abriga as sedes de importantes bancos e instituições financeiras, enquanto C2 é uma área comercial com muitas lojas, um importante terminal de metrô e também diversos pontos turísticos. Tanto C1 quanto C2 têm um grande fluxo de pessoas durante todo o ano.

As áreas residenciais R1 e R2 diferem em termos do nível econômico dos moradores: R1 é um bairro de classe média, enquanto R2 é uma área mais rica, com edifícios e casas luxuosas.

A Fig. 6, canto inferior direito, representa a região C1, selecionada desenhando uma polilinha ao longo da avenida principal do distrito financeiro (T1), e destaca a vista do tipo radial (C1-c) dos três tipos de crimes mais prevalentes de dois locais em C1 (indicados pelas setas). A visualização do tipo de classificação (C1-d) na parte inferior mostra como a incidência dos cinco crimes mais frequentes varia ao longo do tempo. Dois tipos de crime lideram o ranking ao longo dos anos (as curvas bege e rosa no topo): roubo a pedestres e arrombamento de veículos. Analisando a visão radial do tipo (C1-b e C1-c),

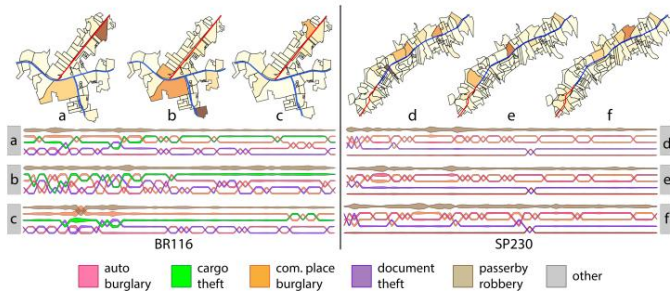


Fig. 7. Pontos críticos ao redor das rodovias BR116 e SP230 considerando todos os tipos de crimes.

Dos locais destacados, pode-se notar que esses dois tipos de crime são de fato os predominantes nessas regiões (codificados pela cor). Inspeccionando outros locais simplesmente clicando neles na visualização do mapa, concluímos que roubo a pedestres e arrombamento de veículos são os tipos de crime predominantes em quase todos os locais em C1.

É importante, no entanto, que esses achados sejam interpretados no contexto da hipótese de que a distribuição espacial de roubos e furtos de veículos por pedestres é moldada pela configuração da malha viária. Observe que a malha viária em C1 é a mais regular; podemos identificar isso pelo número de quarteirões bem definidos na área de análise. Um quarteirão com essa característica é comum em áreas urbanas mais consolidadas e centrais, o que nos leva a conjecturar uma relação entre infraestrutura urbana e risco de roubo.

Realizando a mesma análise na região C2 (canto superior direito da Fig. 6), que foi selecionada clicando e expandindo o local central da região (a marrom), observamos um comportamento diferente.

A visualização por tipo de classificação (C2-d) mostra que há um tipo de crime que cresceu ao longo dos anos (curva verde): o roubo de carga. Selecionando o roubo de carga na visualização radial na interface do CrimAnalyzer (Fig. 5(h)), a visualização em mapa (Fig. 5(b)) revela que o roubo de carga não é prevalente em toda a região, mas está concentrado em apenas alguns locais, sendo o local marrom-escuro no centro da região. Observe que o roubo de carga se tornou o terceiro tipo de crime mais comum naquela região ao longo do tempo, ficando atrás apenas do roubo a pedestres e do roubo de documentos. Outros locais apresentam um comportamento mais uniforme, tendo roubo a pedestres, roubo de veículos e roubo a estabelecimentos comerciais como os principais tipos de crime.

Movendo-se do centro da cidade para áreas mais residenciais, a análise revela uma mudança substancial nos padrões de criminalidade, como se pode observar à esquerda da Figura 6, onde o padrão de criminalidade em R1 e R2 é resumido. Na região residencial R1, por exemplo (canto superior esquerdo da Figura 6), o roubo a pedestres continua sendo o tipo de crime mais comum, seguido pelo roubo de documentos. No entanto, alguns locais em R1 têm o roubo a ônibus (passageiros e/ou motoristas de ônibus públicos são roubados) como a segunda atividade criminosa mais comum (R1-d). O local laranja apontado pela seta superior é um exemplo (R1-b). A análise de padrões de criminalidade local por local é fácil de realizar com o CrimAnalyzer, neste caso, já que o número de locais é pequeno e os usuários precisam apenas selecionar o local no mapa para que seu padrão de criminalidade seja revelado. A importância do roubo a ônibus em R1 é facilmente percebida na visualização do tipo de classificação (R1-d) representada na parte inferior, onde a curva azul (roubo a ônibus) atinge níveis de classificação elevados em diversas oportunidades.

Semelhantemente ao que acontece em C1, C2 e R1 (e também na maior parte da cidade), a região R2 (embaixo à esquerda na Fig. 6) tem o roubo a pedestres como o tipo de crime predominante, o que pode ser claramente

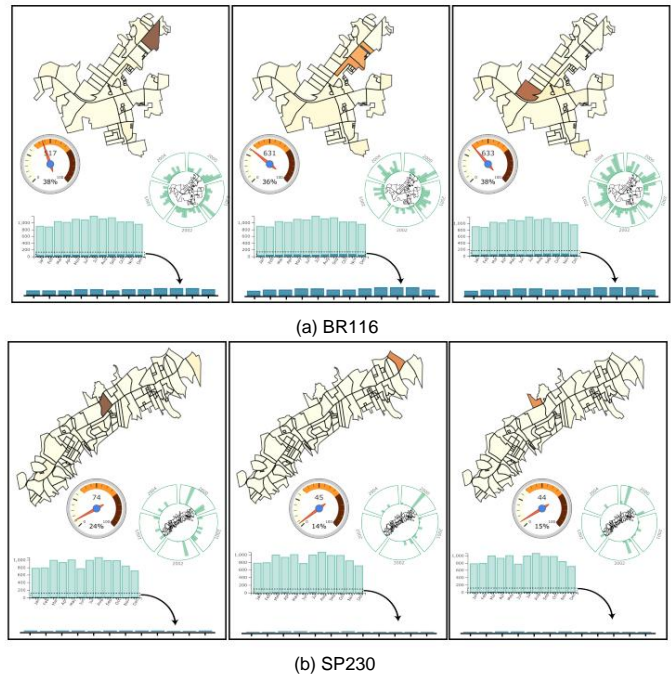


Fig. 8. Pontos críticos de roubo de carga ao longo de duas importantes rodovias, BR116 e SP230. A BR116 apresenta um número muito maior e mais frequente de roubos de carga do que a SP230.

Visto a partir da perspectiva do tipo de classificação (R2-d). No entanto, os padrões de criminalidade variam consideravelmente entre os locais, e alguns deles nem sequer têm o roubo a pedestres como crime prevalente, como os dois locais destacados, que têm o roubo a pedestres como o segundo em importância (R2-c e R2-d). Além disso, o roubo a residências é o crime mais relevante em uma dessas regiões. De fato, o roubo a residências é um crime relevante em R2 como um todo, como indicado pela curva avermelhada na perspectiva do tipo de classificação (R2-d). Observe que o roubo a residências aumentou em importância ao longo dos anos.

A discussão acima mostra que as funcionalidades de análise visual implementadas no CrimAnalyzer são capazes de classificar as tarefas analíticas T1, T2 e T3 de forma simples, intuitiva e eficaz. A flexibilidade para lidar com bairros espacialmente complexos com diferentes formatos permite aos usuários examinar conjuntos de quarteirões, bem como regiões ao longo de avenidas e ruas (tarefa analítica T1). A combinação da visualização por classificação e da visualização radial permite aos usuários compreender o padrão de criminalidade em cada região e em locais específicos, evidenciando como os padrões de criminalidade mudam pela cidade e até mesmo de local para local em uma região específica (tarefa analítica T2), uma tarefa difícil de ser realizada sem nossa infraestrutura. Em particular, as visualizações por classificação e por radial demonstram-se eficazes na revelação do comportamento temporal dos padrões de criminalidade, evidenciando que os padrões mudaram ao longo dos anos (tarefa analítica T3). Com os recursos visuais disponíveis, essa análise seria um processo árduo, exigindo a implementação de múltiplos filtros e ferramentas numéricas e computacionais sofisticadas. De fato, a dificuldade em realizar uma análise semelhante com os sistemas analíticos existentes se deve, em parte, às limitações de seus recursos visuais e, em parte, à inadequação das ferramentas existentes para revelar informações essenciais ocultas nos dados.

6.2 Análise de Hotspot e Roubo de Carga (T1, T4, T5)

Este estudo de caso foi conduzido por especialistas da área, e eles estavam interessados em um tipo específico de crime: o roubo de carga. Embora o roubo de carga não esteja entre os crimes mais proeminentes,

PARA APARECER NO IEEE TVCG

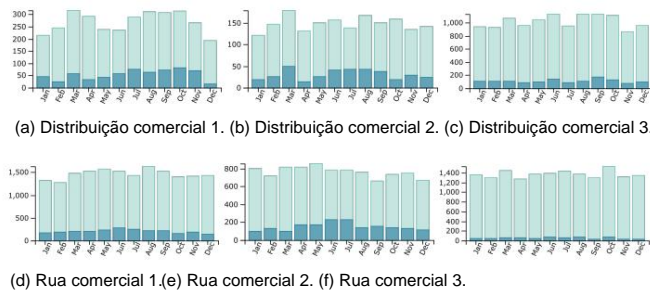


Fig. 9. Os assaltos a estabelecimentos comerciais tendem a aumentar durante o inverno (o inverno na América do Sul vai de meados de junho a meados de setembro).

A análise dos tipos de crimes em São Paulo é de grande interesse devido às suas características espaciais, aos altos valores envolvidos e à participação de quadrilhas violentas nesse tipo de atividade criminosa. É sabido que roubos (ou furtos) de cargas de alto valor tendem a ocorrer perto das principais rodovias que ligam São Paulo a outras regiões do Brasil. Portanto, especialistas da área concentraram sua análise em duas importantes rodovias: a SP230, que liga São Paulo a estados do sul do Brasil, e a BR116, que liga São Paulo ao Rio de Janeiro.

Para realizar sua análise, os especialistas de domínio utilizaram a ferramenta de seleção de polilinhas para selecionar um número considerável de locais ao longo das rodovias e avenidas que conectam a cidade às rodovias. O número de regiões envolvidas nessas análises torna a investigação local por local tediosa, tornando os hotspots uma alternativa melhor. A Figura 7 mostra três hotspots obtidos das regiões selecionadas ao longo da BR116 e SP230 e avenidas próximas. As rodovias são destacadas em vermelho e as avenidas próximas em azul nos mapas de hotspots representados na Figura 7. A visualização do tipo de classificação revela os padrões de criminalidade em cada hotspot (considerando apenas os cinco tipos de crime mais relevantes). Observe que na BR116, o roubo de carga figura entre os crimes mais relevantes (linhas verdes), tornando-se o segundo crime mais relevante em vários momentos. Na SP230, no entanto, o roubo de carga não é predominante, não aparecendo entre os cinco crimes mais relevantes na visualização do tipo de classificação em nenhum hotspot.

No SP230, o padrão predominante é roubo a pedestres, arrombamentos de veículos e arrombamentos de estabelecimentos comerciais. O CrimAnalyzer esclarece quais locais são relevantes em cada hotspot, seus padrões de criminalidade e como os padrões de criminalidade evoluem, abordando adequadamente as tarefas analíticas T4 e T5.

No entanto, os especialistas estão interessados em roubo de carga. Para centralizar a análise em um único tipo de crime, os usuários precisam apenas selecionar esse tipo na visualização radial, filtrando os dados de forma que os hotspots e todas as visualizações sejam atualizados para representar apenas informações relacionadas ao tipo de crime selecionado. A Figura 8 mostra os hotspots associados apenas ao roubo de carga. Os widgets de medidor mostram que o número de roubos de carga na BR116 é uma ordem de magnitude maior do que no SP230, apresentando também uma taxa de ocorrência maior. A evolução temporal (visualização radial) no centro-direita de cada grade mostra o comportamento temporal do roubo de carga em cada hotspot. É claro que o número de roubos de carga no SP230 diminuiu ao longo dos anos, enquanto no BR116 nenhuma redução é observada. Os histogramas abaixo do widget de medidor mostram a intensidade do roubo de carga (as barras escuras curtas) em cada mês, comparando-os com o número total de crimes na região.

As ferramentas de visualização do CrimAnalyzer também deixam claro que, na BR116, o roubo de cargas ocorre principalmente ao longo da rodovia (curvas vermelhas nos mapas da BR116 na Fig. 7), enquanto na SP230 os locais relevantes de cada hotspot estão localizados na avenida que liga a BR116 à BR116.

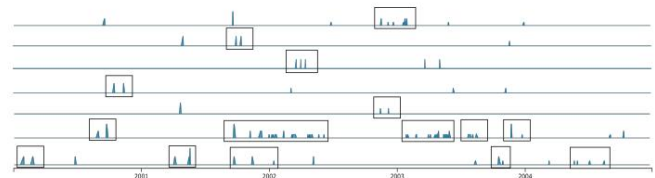


Fig. 10. O fenômeno da quase repetição de vitimização. Quando uma casa é assaltada, o risco de reincidência em um curto período de tempo não é apenas maior para a casa-alvo, mas também para as casas vizinhas.

rodovia para a cidade (curvas azuis nos mapas da SP230 na Fig. 7). Especialistas do domínio consideraram esta uma descoberta importante porque é sabido que o modus operandi dos criminosos e, consequentemente, a localização do roubo de carga muda de acordo com o produto transportado. Portanto, a possibilidade de identificar essas estradas deve tornar as políticas de segurança pública mais eficientes. Outro aspecto interessante apontado pelos especialistas é a capacidade de revelar hotspots associados a atividades criminosas esparsas, como o representado na Fig. 8(b) (veja os picos na vista radial). Hotspots esparsos são relevantes e merecem ser investigados, pois podem estar associados a características locais que provavelmente aumentariam a chance de crimes serem cometidos. Observe que essas descobertas dificilmente poderiam ser feitas sem os recursos visuais habilitados por sistemas como o CrimAnalyzer.

6.3 Sazonalidade e o Elemento Temporal do Crime (T3)

Este estudo de caso corrobora se alguns comportamentos criminosos descritos e validados em trabalhos anteriores também ocorrem em São Paulo.

Sazonalidade Um aspecto importante relacionado às atividades criminosas é a sazonalidade. Há uma série de estudos na literatura que apoiam a hipótese de que certos tipos de crimes são sazonais, enquanto outros não. Por exemplo, van Koppen e Jansen [50] argumentam que, na Holanda, o roubo em estabelecimentos comerciais aumenta durante o inverno devido ao aumento do número de horas escuras durante o dia. Na América do Sul, o inverno geralmente começa em meados de junho e dura até meados de setembro, durante este período, especialmente em julho e agosto, o número de horas escuras é maior do que no resto do ano. Uma questão interessante relacionada à tarefa T3 é se as descobertas de van Koppen e Jansen são válidas em São Paulo. Para procurar uma resposta, confiamos no CrimAnalyzer para explorar seis grandes áreas comerciais na cidade de São Paulo, três distritos comerciais e três ruas comerciais populares. A Fig. 9 mostra a visão temporal cumulativa de cada uma das regiões analisadas.

Os histogramas mais escuros sobrepostos correspondem ao número de roubos e furtos em estabelecimentos comerciais em cada mês. O histograma sobreposto é gerado simplesmente selecionando "roubo em estabelecimentos comerciais" na visualização por tipo temporal.

A Figura 9 mostra claramente que cinco das seis regiões apresentam um aumento no número de roubos e furtos em estabelecimentos comerciais durante o inverno (ae), corroborando as conclusões de van Koppen e Jansen. Embora não possamos afirmar com certeza que a hipótese seja verdadeira, a análise possibilitada pelo CrimAnalyzer fornece evidências sobre a sazonalidade desse tipo de crime, ajudando assim a responder a uma das perguntas associadas à tarefa T3.

A teoria **da quase repetição da vitimização** afirma que quando uma casa é assaltada, o risco de reincidência é

não apenas maior para o domicílio alvo, mas também para os domicílios próximos, com período de risco que parece decair após algumas semanas ou meses [37]. A teoria da quase repetição da vitimização encontrou evidências de sua veracidade em vários países, mas não encontramos nenhum relato sobre ela em São Paulo.

Utilizando o CrimAnalyzer, examinamos duas regiões de São Paulo onde o roubo residencial é um crime recorrente, incluindo a região R2 discutida no estudo de caso apresentado na Seção 6.1. A Figura 10 mostra a série temporal, em escala diária, de sete locais nas regiões analisadas, que varia em termos de frequência de crimes e número de roubos residenciais. Os picos em caixa indicam eventos de roubo residencial que ocorrem com menos de trinta dias de intervalo. Observe que, mesmo em locais onde o roubo residencial é realmente ocasional (linhas 2 a 5 na Figura 10), o fenômeno de vitimização quase repetida pode ser claramente observado.

A sazonalidade e a quase repetição da vitimização são facilmente observadas com o CrimAnalyzer, possibilitando uma série de possibilidades analíticas. Por exemplo, em estações mais quentes, a luz do dia dura mais, incentivando um número maior de pessoas a permanecer nas ruas, aumentando sua exposição a atos ilícitos e atividades criminosas.

Durante as festas de fim de ano, é comum que pessoas viajem para o interior, deixando seus pertences desprotegidos, o que facilita assaltos e outras formas de crime. Esses fenômenos também podem ser analisados com o CrimAnalyzer.

7 AVALIAÇÃO DOS ESPECIALISTAS

Depois de usar o CrimAnalyzer e executar uma variedade de experimentos, incluindo o estudo de caso relatado na Seção 6.2, os especialistas do domínio nos deram o seguinte feedback.

Apesar de suas limitações, o CrimAnalyzer nos permitiu compreender melhor os desafios ainda não elucidados pelas ferramentas convencionais de análise criminal. Primeiro, ao utilizar recursos matemáticos e computacionais sólidos para revelar atividades criminosas georreferenciadas, o CrimAnalyzer incita a busca por explicações plausíveis para os padrões criminais observados, o que seria impossível com a análise convencional. Segundo, o CrimAnalyzer motiva a reflexão sobre a relação entre os diferentes tipos de crime e sobre conexões topológicas, direcionais e relacionais que podem afetar o número de crimes em locais e intervalos de tempo específicos.

Em terceiro lugar, uma ferramenta analítica que permite a análise de crimes em locais específicos leva a pensar a cidade em sua complexidade e, ao mesmo tempo, orienta a investigação das características urbanas (administrativas, demográficas, físicas e sociais) e sua interação, da qual resultam os padrões locais observados. Em quarto lugar, o CrimAnalyzer revela a heterogeneidade da cidade quanto à sua infraestrutura urbana, as diferenças entre áreas comerciais, financeiras e residenciais, o fluxo de pessoas, o transporte público e privado, bem como a necessidade de melhorias, não apenas em termos de policiamento em locais específicos e de acordo com o tipo de crime, mas também, e principalmente, em termos de ferramentas para auxiliar a investigação criminal a fim de reduzir os altos índices de impunidade. Finalmente, em contraste com a metodologia estatística mais simplista, a abordagem determinística para identificação de hotspots torna-se fundamental para enfatizar a dinâmica dos processos espaço-temporais e capturar manifestações sociais típicas, como crimes.

Os especialistas ficaram bastante entusiasmados com o CrimAnalyzer, pois permitiu que compreendessem e levantassem hipóteses sobre uma série de fenômenos, como no caso do roubo de carga, que seriam difíceis de outra forma. Especificamente, um dos especialistas afirmou: “Analisando a vasta quantidade de informações disponibilizada pelo CrimAnalyzer, conseguimos detectar

padrões e tendências espaço-temporais que nos permitirão melhorar as políticas públicas...”.

8 DISCUSSÃO E LIMITAÇÕES

O CrimAnalyzer foi desenvolvido em estreita colaboração com especialistas da área. A versão atual atende aos requisitos deles, porém, algumas limitações e trabalhos futuros foram identificados como parte da nossa colaboração.

Estabilidade da NMF. Nossa abordagem para identificar pontos críticos não é estável, pois a técnica de Fatoração de Matrizes Não Negativas depende das condições iniciais do procedimento de otimização. Para neutralizar esse efeito, algumas implementações, como a que estamos usando em nosso sistema, nos permitem executar o método diversas vezes, mantendo a solução com o menor erro. Embora os resultados se tornem bastante estáveis após habilitar a alternativa de múltiplas execuções, uma abordagem mais robusta poderia ser buscada para mitigar possíveis efeitos.

Discretização Espacial. A discretização espacial utilizada no CrimAnalyzer são as unidades censitárias de São Paulo. Adotamos essa medida porque nossos colaboradores tinham interesse em ver a análise nesse nível de detalhe. No entanto, estamos cientes do problema da unidade de área modificável (MAUP), ou seja, as unidades censitárias não representam "unidades naturais" de análise e o resultado de determinadas análises pode mudar ao modificar a unidade de agregação [8]. Um trabalho futuro imediato seria estender e flexibilizar nossa discretização espacial. Dessa forma, poderemos aplicar nossa ferramenta em outros cenários.

Múltiplas fontes de dados. Eventos criminais por si só raramente contam a história toda. Dados adicionais podem ser usados para aprimorar a compreensão da camada criminal. Por exemplo, a presença de bares e pubs, distância de parques, terrenos e edifícios baldios, clima, entre outras informações, podem ter relação com certas atividades criminosas. Dado o crescente número de iniciativas para tornar os dados publicamente disponíveis, estamos considerando combinar essas informações para aprofundar o entendimento dos crimes em áreas urbanas. Uma ferramenta matemática interessante nesse contexto é a decomposição tensorial, uma generalização da decomposição matricial capaz de extrair padrões de múltiplas fontes de dados. O desenvolvimento de ferramentas analíticas visuais para mapear informações de decomposição tensorial em conteúdo visual é um problema importante [6] que mal foi abordado no contexto da análise criminal.

Abordagem global vs. local. O CrimAnalyzer utiliza uma abordagem local para explorar e analisar padrões de criminalidade. Embora essa fosse uma exigência dos especialistas da área, e concordemos que era a abordagem correta para o problema, principalmente porque os especialistas da área possuem conhecimento prévio e hipóteses sobre comportamentos criminosos em locais específicos, em algumas de nossas entrevistas com especialistas da área, discutimos a opção de ter uma técnica global que pudesse processar todo o espaço e propor locais interessantes para serem explorados. Essa alternativa foi aceita pelos especialistas, mas como uma técnica complementar. Como trabalho futuro, também estamos interessados em abordar esse problema de ambas as perspectivas (global e local).

Várias cidades e cenários diferentes Por fim, pretendemos aplicar e validar nosso sistema em outras cidades e países. Atualmente, estamos coletando dados criminais de diversas localidades e, em breve, esperamos lançar o sistema para analisar diversas cidades no Brasil. Além disso, nossa abordagem pode ser estendida a outros cenários além da análise criminal.

Por exemplo, pode-se usar o sistema para analisar a dinâmica dos acidentes de trânsito em locais específicos da cidade, possibilitando

para descobrir como o número de acidentes entre carros, entre carros e ônibus, atropelamentos , etc. evoluem ao longo do tempo.

9 CONCLUSÃO

Implementamos uma ferramenta de análise visual para auxiliar na análise de crimes em regiões locais. Desenvolvemos o CrimAnalyzer em estreita colaboração com especialistas da área e traduzimos suas análises para o sistema de visualização. Também propomos uma técnica baseada em NMF para identificar pontos críticos. Nosso sistema foi validado por comparações qualitativas e quantitativas, além de estudos de caso utilizando dados reais e com o feedback dos especialistas da área. Além disso, verificamos dois comportamentos criminais (sazonalidade e quase repetição de vitimização) utilizando dados criminais de São Paulo.

AGRADECIMENTOS

Este trabalho foi apoiado pelo CNPq-Brasil (processos n.º 302643/2013-3 e n.º 301642/2017-63), CAPES-Brasil (processos n.º 10242771) e Fundação de Amparo à Pesquisã do Estado de São Paulo (FAPESP)-Brasil (processos n.º 2014/12236-1, n.º 2016/04391-2 e n.º 2017/05416-1). As opiniões expressas são dos autores e não refletem a política ou posição oficial da Fundação de Amparo à Pesquisa do Estãdo de São Paulo. Agradecemos também à Intel por disponibilizar parte dos recursos computacionais que utilizamos no desenvolvimento deste trabalho. Silva é financiado em parte por: Moore-Sloan Data Science Environment da NYU; NASA; NSF concede os prêmios CNS-1229185, CCF-1533564, CNS-1544753, CNS-1730396, CNS-1828576 e DARPA. Quaisquer opiniões, descobertas, conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não refletem necessariamente as opiniões da DARPA.

REFERÊNCIAS

[1] S. Adorno. Democracia em progresso no Brasil contemporâneo: corrupção, crime organizado, violência e novos caminhos para o Estado de Direito. *Revista Internacional de Criminologia e Sociologia*, 2:409–425, 2013.

[2] T. Aljrees, D. Shi, D. Windridge e W. Wong. Identificação de padrões criminais com base em agrupamento k-means modificado. Em 2016, *Conferência Internacional sobre Aprendizado de Máquina e Cibernética (ICMLC)*, vol. 2, pp. 799–806, 2016.

[3] MA Andresen. Mapeando a prevenção do crime: o que fazemos e para onde precisamos ir. Em *Prevenção do Crime no Século XXI*, pp. 113–126. Springer, 2017.

[4] GL Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. v. Landesberger, P. Bak e D. Keim. Mapas auto-organizados de espaço-tempo e tempo-espaço para explorar padrões espaço-temporais. *Fórum de Computação Gráfica*, 29(3):913–922, 2010.

[5] L. Anselin. Indicadores locais de associação espacial – LISA. *Análise Geográfica*, 27:93 – 115, 1995.

[6] R. Ballester-Ripoll, P. Lindstrom e R. Pajarola. Tthresh: Compressão tensorial para dados visuais multidimensionais. *IEEE TVCG*, 2019.

[7] C. Brunsdon, J. Corcoran e G. Higgs. Visualização de espaço e tempo em padrões de criminalidade: comparação de métodos. *Computadores, Meio Ambiente e Sistemas Urbanos*, 31(1):52–75, 2007.

[8] S. Chainey e J. Ratcliffe. *SIG e mapeamento de crimes*. Wiley, 2005.

[9] S. Chainey, S. Reid e N. Stuart. Quando um hotspot é considerado um hotspot? Um procedimento para a criação de mapas de hotspots de criminalidade estatisticamente robustos. Em *Aplicações Sócio-Econômicas da Ciência da Informação Geográfica*, pp. 21–36. Taylor e Francis, 2002.

[10] S. Chainey, L. Thompson e S. Uhlig. A utilidade do mapeamento de hotspots para prever padrões espaciais de crime. *Security Journal*, 21(1):4–28, 2008.

[11] H. Chen, H. Atabakhsh, T. Petersen, J. Schroeder, T. Buetow, L. Chaboya, C. O’Toole, M. Chau, T. Cushman, D. Casey e Z. Huang. COPLINK: Visualização para análise de crimes. Em *Pesquisa de Governo Digital da Conferência Nacional* , 2003.

[12] H. Chen, W. Chung, JJ Xu, G. Wang, Y. Qin e M. Chau. Mineração de dados criminais: uma estrutura geral e alguns exemplos. *Computador*, 37(4):50–56, 2004.

[13] SN de Melo, DVS Pereira, MA Andresen e LF Matias. *Variações espaciais/temporais do crime: uma perspectiva da teoria da atividade rotineira*. *Revista Internacional de Terapia de Ofensores e Criminologia Comparada*, pp. 1–26, 2017.

[14] J. Eck, S. Chainey, J. Cameron e R. Wilson. Mapeamento do crime: compreensão dos pontos críticos. *Relatório técnico*, Instituto Nacional de Justiça, 2005.

[15] E. Eftelioglu, S. Shekhar e X. Tang. Detecção de pontos críticos de crimes: uma perspectiva computacional, pp. 82–111. IGI Global, 2016.

[16] R. Gao, H. Tao, H. Chen, W. Wang e J. Zhang. Design de visualização coordenada de exibição múltipla para análise de resolução de crimes: Grande desafio 2014: Menção honrosa para uso eficaz de visualizações coordenadas. *Em Conf. VAST*, pp. 321–322, 2014.

[17] A. Getis e JK Ord. A análise da associação espacial pelo uso de estatísticas de distância. *Análise Geográfica*, 24(3):189–206, 1992.

[18] D. Guo, J. Chen, AM MacEachren e K. Liao. Um sistema de visualização para padrões espaço-temporais e multivariados. *IEEE TVCG*, 12(6):1461–1474, 2006.

[19] D. Guo e J. Wu. Compreendendo padrões espaço-temporais de múltiplos tipos de crimes com uma abordagem de análise geovisual. Em *Modelagem e Mapeamento de Crimes Usando Tecnologias Geoespaciais*, pp. 367–385. Springer, 2013.

[20] J. Hagenauer, M. Helbich e M. Leitner. Visualização de trajetórias criminais com mapas auto-organizáveis: um estudo de caso sobre a avaliação do impacto de furacões em pontos críticos de criminalidade espaço-temporal. Em *Anais da 25ª conferência da Associação Cartográfica Internacional*, 2011.

[21] T. Hart e P. Zandbergen. Estimativa de densidade de kernel e mapeamento de hotspots: examinando a influência do método de interpolação, tamanho da célula da grade e largura de banda na previsão de crimes. *Policing: An International Journal*, 37(2):305–323, 2014.

[22] E. Johansson, C. Gahlin e Å. Borg. Pontos críticos de criminalidade: uma avaliação da técnica de mapeamento espacial KDE. *Conferência Europeia de Informática de Inteligência e Segurança de 2015*, pp. 69–74, 2015.

[23] R. Kerry, P. Goovaerts, RP Haining e V. Ceccato. Aplicação de análise geoestatística a dados criminais: roubos de automóveis nos estados bálticos. *Análise Geográfica*, 42(1):53–77, 2010.

[24] H. Kim e H. Park. Fatorações de matrizes esparsas não negativas por meio de mínimos quadrados alternados com restrição de não negatividade para análise de dados de microarray. *Bioinformática*, 23(12):1495–1502, 2007.

[25] M. Kuldorff. Uma estatística de varredura espacial. *Comunicações em Estatística-Teoria e métodos*, 26(6):1481–1496, 1997.

[26] J. Lampinen e T. Kostiainen. Modelo de densidade de probabilidade generativa no mapa auto-organizado. Em *Redes Neurais Auto-organizadas: Avanços e Aplicações Recentes*, pp. 75–94. Springer, 2002.

[27] DD Lee e HS Seung. Algoritmos para fatoração de matrizes não negativas . Em *Advances in Neural Information Processing Systems*, pp. 556–562, 2001.

[28] N. Levine. *CrimeStat IV: Um programa de estatística espacial para a análise de locais de incidentes criminais*. *Relatório técnico*, Instituto Nacional de Justiça, 2013.

[29] D. Li, Y. Wang, S. Wu, J. Qi e T. Wang. Uma abordagem de análise visual para explorar padrões criminais com base em dados multidimensionais. Em 2017, *IEEE International Geoscience and Remote Sensing Symposium*, pp. 5563–5566, 2017.

[30] J. Lukasczyk, R. Maciejewski, C. Garth e H. Hagen. Compreendendo hotspots: Uma abordagem de análise visual topológica. Em *Anais da 23ª Conferência Internacional SIGSPATIAL sobre Avanços em Sistemas de Informação Geográfica*, pp. 36:1–36:10, 2015.

[31] T. Munzner. Um modelo aninhado para design e validação de visualização. *IEEE TVCG*, 15(6):921–928, 2009.

[32] T. Nakaya e K. Yano. Visualização de aglomerados de crimes em um cubo espaço-temporal: Uma abordagem exploratória de análise de dados usando estimativa de densidade de kernel espaço-temporal e estatísticas de varredura. *Transactions in GIS*, 14(3):223–239, 2010.

[33] MA Oliver e R. Webster. Krigagem: um método de interpolação para sistemas de informação geográfica. *International Journal of Geographical Information System*, 4(3):313–332, 1990.

[34] JK Ord e A. Getis. Estatísticas de autocorrelação espacial local: questões distribucionais e uma aplicação. *Análise geográfica*, 27(4):286–306, 1995.

[35] DW Osgood. Modelos estatísticos de eventos da vida e comportamento criminoso. Em *Handbook of Quantitative Criminology*, pp. 375–396. Springer, 2010.

[36] N. Otsu. Um método de seleção de limiar a partir de histogramas de níveis de cinza. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[37] N. Polvi, T. Looman, C. Humphries e K. Pease. A evolução temporal da vitimização por roubo recorrente. *The British Journal of Criminology*, 31(4):411–414, 1991.

[38] J. Ratcliffe. Mapeamento do crime: desafios espaciais e temporais. Em *Handbook of Quantitative Criminology*, pp. 5–24. Springer, 2010.

- [39] JH Ratcliffe. Análise aorística: a interpretação espacial de eventos temporais inespecíficos. *International Journal of Geographical Information Science*, 14(7):669–679, 2000.
- [40] JH Ratcliffe. Assinaturas aorísticas e a análise espaço-temporal de padrões de criminalidade de alto volume. *Journal of Quantitative Criminology*, 18(1):23–43, 2002.
- [41] JH Ratcliffe. A matriz de hotspot: uma estrutura para a segmentação espaço-temporal da redução da criminalidade. *Prática e Pesquisa Policial*, 5(1):4–28, 2004.
- [42] JH Ratcliffe. Uma teoria de restrição temporal para explicar padrões de infrações espaciais baseados em oportunidades. *Journal of Research in Crime and Delinquency*, 43(3):261–291, 2006.
- [43] SJ Rey e L. Anselin. PySAL: Uma biblioteca Python de métodos analíticos espaciais. *The Review of Regional Studies*, 37(1):5–27, 2007.
- [44] C. S e S. Cha. Um levantamento de medidas binárias de similaridade e distância. *Revista de Sistemática, Cibernética e Informática*, pp. 43–48, 2010.
- [45] RB Santos. Análise de Crime com Mapeamento de Crime. SAGE, 4ª ed., 2016.
- [46] LJS Silva, S. Fiol-Gonzalez, CFP Almeida, SDJ Barbosa e H. Lopes. CrimeVis: Um sistema de visualização interativo para análise de dados criminais no estado do Rio de Janeiro. 19ª Conferência Internacional sobre Sistemas de Informação Empresarial (ICEIS), pp. 193–200, 2017.
- [47] M. Townsley. Visualização de padrões de espaço-tempo no crime: o gráfico de ponto crítico. *Padrões de criminalidade e análise*, 1(1):61–74, 2008.
- [48] M. Townsley. Mapeamento do crime e análise espacial. Em *Prevenção do Crime no Século XXI*, pp. 5–23. Springer, 2017.
- [49] P. Valdivia, F. Dias, F. Petronetto, CT Silva e LG Nonato. Visualização baseada em wavelets de dados variáveis no tempo em gráficos. Em *IEEE VAST*, pp. 1–8, 2015.
- [50] P. van Koppen e R. Jansen. O momento do roubo: variações no tempo do número de roubos comerciais. *Journal of Research in Crime and Delinquency*, 36(1):7–29, 1999.
- [51] D. Wang, W. Ding, H. Lo, M. Morabito, P. Chen, J. Salazar e T. Stepinski. Compreendendo a distribuição espacial do crime com base em suas variáveis relacionadas, utilizando padrões discriminativos geoespaciais. *Computadores, Meio Ambiente e Sistemas Urbanos*, 39:93–106, 2013.
- [52] T. Wang, C. Rudin, D. Wagner e R. Sevieri. Aprendendo a detectar padrões de crime. Em *Aprendizado de Máquina e Descoberta de Conhecimento em Bancos de Dados*, pp. 515–530, 2013.
- [53] J. Xu e H. Chen. Análise e visualização de redes criminosas. *Comun. ACM*, 48(6):100–107, 2005.
- [54] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma e N. Yadav. Detecção de padrões de crime, análise e previsão. Em *Int. Conf. Electr., Communic. Aero. Tecnologia. (ICECA)*, vol. 1, pp. 225–230, 2017.
- [55] Z. Ying. Análise da correlação de fatores de criminalidade com base na tecnologia de mineração de dados. Em *Int. Conf. Robôs Intel. Sys. (ICRIS)*, pp. 103–106, 2016.
- [56] M. Zitnik e B. Zupan. Nimfa: Uma biblioteca Python para fatoração de matrizes não negativas. *Journal of Machine Learning Research*, 13:849–853, 2012.