

Serviço de Sugestão de Rotas Seguras para Veículos

Lucas Zanco Ladeira¹, Allan Mariano de Souza^{1,2}, Geraldo Pereira Rocha Filho³,
Thiago Henrique Silva⁴, Leandro Aparecido Villas¹

¹Laboratório de Redes de Computadores, Universidade Estadual de Campinas,
Cidade Universitária, Campinas, SP, Brasil

²Instituto de Ciência da Computação e Matemática Aplicada,
Universidade de Bern, Suíça

³Departamento de Ciência da Computação, Universidade de Brasília,
Campus Darcy Ribeiro, Brasília, DF, Brasil

⁴Departamento de Informática, Universidade Tecnológica Federal do Paraná
Av. 7 de Setembro, 3165, Curitiba, PR, Brasil

{lucaszl,allanms}@lrc.ic.unicamp.br, geraldof@unb.br,

thiagoh@utfpr.edu.br, leandro@ic.unicamp.br

Abstract. *Considering the increasing number of vehicles in the cities, applications able to inform the traffic situation on the roads become more used. These applications have the objective of suggesting routes, considering congestion and accidents, identified through participatory sensing. However, there are other problems encountered during a course that directly affect the user, such as crime in a particular region. Regions affected by high crime rates may evolve during the day, depending on the type of crime and the density of people in a particular region. Therefore, this work has the objective of proposing a service capable of identifying areas with high criminal incidence, taking into account the evolution of the scenario, and suggesting safe routes that avoid these regions. The proposed solution allows to obtain a safer route without compromising the time spent on the route, in addition, the results found verify the ability to avoid dangerous regions, making safer the route of a driver.*

Resumo. *Considerando o crescente número de veículos nas cidades, aplicações capazes de informar a situação do tráfego nas vias se tornam mais utilizadas. Essas aplicações tem o objetivo de sugerir rotas, considerando congestionamentos e acidentes, identificados através de sensoriamento participativo. No entanto, existem outros problemas encontrados durante um percurso que afetam diretamente o usuário, como por exemplo, a criminalidade em uma determinada região. As regiões afetadas por uma alta incidência criminal podem evoluir durante o dia, dependendo do tipo do crime e da densidade de pessoas em uma determinada região. Portanto, esse trabalho tem o objetivo de propor um serviço capaz de identificar áreas com alta incidência criminal, levando em consideração a evolução do cenário, e sugerir rotas seguras que evitam essas regiões. A solução proposta permite obter uma rota mais segura sem comprometer o tempo do percurso, além disso, os resultados encontrados proporcionam verificar a capacidade de evitar regiões perigosas, tornando mais seguro o percurso de um condutor.*

1. Introdução

O crescimento demográfico nas cidades aumenta o fluxo de veículos, tornando mais comum a ocorrência de congestionamentos, diminuindo a segurança nas vias, e aumentando a emissão de CO_2 na atmosfera. Existem várias aplicações que tentam colaborar com o gerenciamento eficiente do tráfego, tais como: sincronização de semáforos, monitoramento em tempo real de ônibus, identificação de vias congestionadas, entre outras. Essas aplicações compõem uma categoria de sistemas chamada Sistema de Transporte Inteligente (ITS). De acordo com [Cunha et al. 2018], ITSs buscam resolver problemas ligados ao transporte nas cidades, e para isso, utilizam dados referentes ao tráfego, comunicação entre dispositivos, e computação. Uma categoria de ITSs é focada na identificação de congestionamentos e controle do fluxo de veículos, e os sistemas que compõem essa categoria são conhecidos como Sistemas de Gerenciamento de Tráfego (TMS) [de Souza and Villas 2016]. Sistemas de sugestão de rotas também fazem parte dessa categoria, e utilizam diferentes estratégias para identificar áreas congestionadas, com a finalidade de sugerir rotas alternativas que diminuem o tempo de percurso.

No entanto, outras informações podem ser utilizadas para identificar possíveis rotas para usuários. Como exemplo, a criminalidade que afeta diretamente o condutor durante um percurso, como pode ser observado em alguns casos de morte pela escolha de rotas mais rápidas [Farias 2018]. Para tal, podem ser identificadas as regiões com alta incidência criminal, as quais devem ser evitadas. Um problema encontrado na maior parte dos trabalhos que consideram os dados contextuais, é a falta da análise no impacto da evolução do cenário para verificar as mudanças que ocorrem [Kim et al. 2014, Quercia et al. 2014, Quercia et al. 2015, Santos et al. 2018a, de Souza et al. 2018]. A evolução do cenário é caracterizada nesse trabalho por uma janela temporal. De acordo com [ArcGIS 2018], uma janela temporal é um período de tempo entre um início e um fim, ou seja, um intervalo de tempo. Sendo assim, é possível identificar um intervalo de tempo que as informações criminais representam o cenário criminal das regiões. Se tratando dos trabalhos que utilizam dados criminais, muitos deles consideram apenas a densidade e não o tipo do crime na identificação das regiões com alta incidência criminal [Galbrun et al. 2016, Utamima and Djunaidy 2017].

Considerando esse conceito, o objetivo desse trabalho é de propor um serviço de sugestão de rotas, ciente da evolução do cenário, capaz de representar a segurança de cada rota de forma probabilística. O serviço é dividido em 3 módulos distintos, compreendendo: Mineração, capaz de obter as informações para a seleção das rotas, Clusterização, identifica regiões com alta incidência criminal, e Sugestão, calcula a segurança probabilística de cada rota e sugere a melhor. Baseando-se na segurança das rotas é possível identificar diferentes métricas capazes de descrever o cenário criminal durante meses distintos. Os resultados obtidos apresentam a capacidade do serviço de identificar rotas seguras e as distinguir de outras rotas possíveis.

O trabalho está organizado da seguinte maneira: a Seção 2 apresenta os trabalhos relacionados, com o intuito de distinguir os que utilizam dados criminais dos que avaliam a evolução do cenário, e apresentar alguns problemas de cada trabalho. Na Seção 3, são apresentadas as definições utilizadas no trabalho. A solução proposta é apresentada na Seção 4. Na Seção 5 são apresentados os resultados obtidos. Por fim, na Seção 6 são apresentadas as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

É explorado extensamente na literatura, a escolha de rotas baseada apenas no menor tempo de percurso [de Souza and Villas 2016, Brennand et al. 2017, Akabane et al. 2017, de Souza et al. 2016, Pan et al. 2017], podendo ser utilizados algoritmos como *k-Shortest Paths* ou A^* para tal tarefa. No entanto, outros dados podem ser utilizados para tornar a escolha da rota mais robusta. Isso se refere a possibilidade de outros fatores acarretarem em um descontentamento por parte de um motorista em uma dada via. Sendo assim, dados sobre a poluição do local [Quercia et al. 2015], sentimento de indivíduos [Quercia et al. 2014], [Santos et al. 2018b], segurança [Shah et al. 2011, Kim et al. 2014, Keler and Mazimpaka 2016, Galbrun et al. 2016, Utamima and Djunaidy 2017, Johnson et al. 2017, Taha 2017, de Souza et al. 2018, Santos et al. 2018a], entre outros, possuem relevância nesse cenário. Como o foco do trabalho é na seleção de rotas seguras, apenas serão explorados mais a fundo os trabalhos que tratam do mesmo cenário.

Nos trabalhos de [Shah et al. 2011, Kim et al. 2014, Utamima and Djunaidy 2017] são propostos sistemas para sugestão de rotas seguras. Esses sistemas utilizam algoritmos de agrupamento para identificar regiões com alta incidência criminal. Alguns problemas encontrados em cada trabalho representam: em [Shah et al. 2011], não é considerado o tráfego dos veículos nas rotas, ou seja, mesmo escolhendo a rota mais curta não há garantia de que seja a mais rápida, em [Kim et al. 2014], as rotas são alteradas adicionando *waypoints* que contornam regiões com um foco de sentimento negativo, portanto, não é possível ter certeza do tempo de percurso de cada via, e em [Utamima and Djunaidy 2017], o tipo do crime serve apenas para informar o usuário, não sendo aplicado filtro algum. Em [Johnson et al. 2017] é proposta uma plataforma para sugestão de rotas alternativas, a qual compara diferentes dados contextuais, como por exemplo, locais inseguros, locais bonitos (parques e museus), entre outros. É possível encontrar o mesmo problema citado no trabalho de [Utamima and Djunaidy 2017], o qual se refere a não utilização da tipificação criminal.

Para solucionar alguns problemas citados anteriormente, *itsSAFE* é proposto por [de Souza et al. 2018]. Ele é um sistema de transporte inteligente, que sugere rotas de acordo com conhecimento sobre o tráfego e a segurança das vias. A situação do tráfego nas vias é identificada através da comunicação entre veículos, pela velocidade média e a posição. O conhecimento sobre a criminalidade é obtido de uma base de dados policial que apresenta os boletins de ocorrência relatados. A seleção das rotas é mapeada como um problema de *Resource Constrained Shortest Path (RCSP)*, e é utilizado um algoritmo de Programação Dinâmica para resolvê-lo em tempo sub-exponencial. Dessa maneira, é identificada a rota mais rápida com segurança menor ou igual a um *threshold*. Nesse trabalho é considerado o cenário no qual uma região é afetada pela criminalidade, e não vias específicas. Todos os trabalhos citados nesse parágrafo e no anterior, não consideram a evolução do cenário criminal, tratando apenas de pontos fixos com histórico de crimes. Ou seja, não é considerada a migração da criminalidade durante o dia, a qual acarreta risco direto aos motoristas.

Um trabalho que considera a evolução do cenário é o de [Santos et al. 2018a], no qual é apresentada uma plataforma para recomendação de rotas baseada em dados contextuais urbanos. Sobre esses dados, é feita uma análise agregada para identificar regiões,

na qual indivíduos possam querer evitar durante um percurso. Mesmo sendo uma plataforma que permite a análise de dados contextuais quaisquer, são utilizados dados sobre crimes vindos de fontes abertas. É aplicada uma clusterização dos dados criminais para identificar regiões com uma alta incidência criminal. Após, dada uma origem e um destino, é calculada a segurança de cada rota considerando a quantidade de clusters que a rota passa, dividido pelo número de clusters encontrados. Por ser uma plataforma genérica, é considerado um tempo de vida dos dados Δt , o qual permite a identificação de uma janela temporal nos dados criminais. No entanto, são considerados clusters circulares, esses podem não representar a disposição das vias sendo possível identificar vias como perigosas sem necessariamente ocorrer crimes próximos a elas. Por fim, a forma que é atribuída a segurança da via não representa a real distribuição dos crimes em cada cluster.

Finalmente, o trabalho de [Galbrun et al. 2016] propõe um algoritmo de sugestão de rotas seguras baseado na mobilidade de pedestres. Para identificar as possíveis rotas, é utilizado um algoritmo de descoberta de rotas sobre mapas extraídos do OpenStreet-Map (OMS). A identificação geográfica da densidade criminal é calculada através de uma *Kernel Density Estimation* (KDE) Gaussiana. Isso possibilita identificar as regiões com alta incidência criminal. Além disso, é considerada uma janela temporal para os dados criminais permitindo representar as mudanças na criminalidade das regiões. A segurança de uma determinada rota, é dada pelo somatório da probabilidade calculada pelo KDE, de todos os vértices da rota. Para escolher a melhor solução, é utilizada uma função bi-objetivo que sugere rotas, considerando a segurança e a distância. Um dos problemas encontrados é que a mobilidade de pedestre é diferente da veicular, na qual é considerado o tráfego. Sendo assim, a rota proposta apenas leva em consideração a distância a ser percorrida. Os algoritmos de descoberta de rotas possuem um alto custo computacional, tornando necessário ignorar parte das soluções, durante a seleção das rotas, para que seja possível utilizá-los em tempo factível.

A solução proposta busca resolver os problemas dos trabalhos citados, de forma a considerar o tipo dos crimes para filtrar a base de dados, representar os crimes em clusters temporais, identificar níveis de segurança distintos dependendo de quão dentro de um cluster uma rota está, e por fim, utilizar o tráfego e o tempo de percurso na escolha das rotas. Por causa dessas limitações, propomos a seguir um novo serviço de sugestão de rotas que permite uma melhor representação do cenário criminal de maneira mais assertiva.

3. Definições Utilizadas

Para desenvolver um serviço de sugestão de rotas seguras, é necessário considerar como é caracterizada a informação sobre os crimes. Isso se refere a tipificação do crime e como identificar quais devem ser considerados durante o roteamento, e, portanto, as regiões que serão evitadas. Um outro aspecto ligado a essa caracterização, é o agrupamento desses crimes, sendo em vias distintas, ou em regiões, conhecidas como clusters. Essas regiões descrevem cada crime de forma a não afetar um ponto específico, mas sim uma região toda. Portanto, existem 3 elementos a serem definidos, um mapa, uma rota, e um cluster.

Um mapa de uma determinada cidade, considerando apenas as vias e suas intersecções, é definido como um grafo dirigido $G = (V, E)$, com um conjunto de vértices $v_i \in V$ (intersecções), e um conjunto de arestas $e_i \in E$ (vias). Cada aresta e_i descreve a

relação entre dois vértices $v_i, v_j \in V \mid v_i \neq v_j$. Cada via possui um atributo de distância δ_i , e condicionados a um estante t , possui um atributo de atraso α_{ti} , e a segurança λ_{ti} . No caso deste trabalho, os vértices também possuem um atributo de segurança μ_{ti} . Uma rota é descrita como um subgrafo $R \subseteq G$, tal que, $R = (V', E')$. Como um cluster se refere a uma região, ele é definido como um grafo não direcionado $C = (VC, EC)$. Como os vértices de G e C podem referenciar a mesma posição geográfica, então pode existir um $v_i \in VC$, e portanto, $vc_i \in V$. A mesma relação pode ser observada entre um cluster e uma rota, pois $R \subseteq G$.

4. Solução Proposta

Nesta seção, é apresentado o serviço de sugestão de rotas proposto, com base na segurança temporal das regiões que os veículos irão percorrer. Na Figura 1, é ilustrado o fluxo de execução do serviço, compreendendo os dados utilizados, cada módulo e a sua organização. O serviço considera dados sobre possíveis rotas, além da criminalidade regional para evitar regiões perigosas. Os dados disponíveis da mobilidade veicular, e a criminalidade, são obtidas no módulo de obtenção e tratamento de dados. Esse módulo tem a função de obter os dados de rotas de veículos considerando o trânsito das vias. As rotas¹ foram obtidas por meio do serviço Here API, armazenadas de maneira estruturada, e utilizadas como entrada para o algoritmo de sugestão. Como a API já permite obter as rotas mais rápidas, não é necessário considerar todas as possíveis rotas para uma mesma origem e destino. Os dados criminais foram obtidos de uma iniciativa de dados abertos que abrange diferentes cidades do mundo.

Após, os dados criminais são utilizados pelo módulo de clusterização. A clusterização busca identificar regiões com alta incidência criminal, com o intuito de que essas regiões sejam evitadas durante um percurso. Para tal, o serviço foi modelado com base no algoritmo DBScan [Santos et al. 2018a] para identificar regiões com alta incidência criminal. Por fim, o módulo de sugestão faz o cálculo da segurança probabilística, correlacionando os clusters encontrados e as rotas mineradas, para identificar a probabilidade da rota estar dentro de cada cluster. Tendo a segurança de cada rota, o serviço proposto seleciona, dentre as rotas mineradas, a melhor opção para sair de uma origem e ir até um destino. A melhor opção se refere a rota mais segura dentre as rotas mais rápidas obtidas pela API. O serviço é dividido nesses 3 módulos, que compreendem cada uma das tarefas descritas, de maneira a possibilitar a simples utilização por Sistemas de Transporte Inteligentes (ITS).

4.1. Clusterização Criminal

Para clusterizar os dados criminais e identificar pontos com alta incidência, é aplicado o algoritmo de clusterização *Density-Based Spatial clustering of applications with noise* (DBScan). Esse é um algoritmo que utiliza a distância entre os pontos mais próximos para identificar os clusters [Santos et al. 2018b]. Ele é geralmente utilizado para encontrar clusters quando o tamanho deles possui grande variação. Essa é uma característica interessante pois, os crimes podem ocorrer concentrados em uma região, ou espalhados na mesma.

¹Podem ser utilizados outros serviços para obter rotas, como por exemplo, a Google Directions API

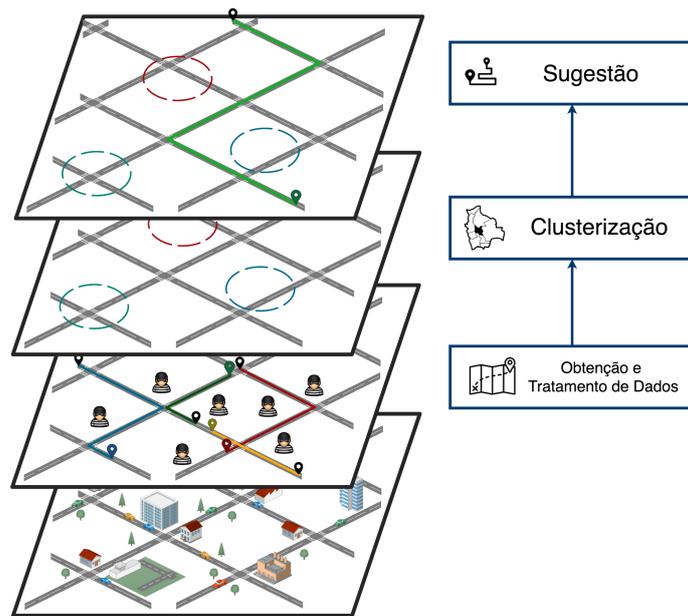


Figura 1. Fluxo da execução do serviço

Para encontrar os parâmetros do DBScan é necessário observar as mudanças nos clusters, considerando tamanho, e distância entre eles. Os parâmetros do DBScan compreendem, o tamanho mínimo da vizinhança v para ser considerado um novo cluster, e a distância máxima ϵ entre dois pontos. Esses dois parâmetros são utilizados para classificar os pontos em 3 tipos distintos, sendo: núcleo, borda, e *outlier*. Os pontos de núcleo possuem uma vizinhança mínima v . Pontos de borda estão a no máximo uma distância ϵ de algum ponto de núcleo. Por fim, os pontos de *outlier*, não estão enquadrados em algum dos requisitos anteriores.

Nesse trabalho, a escolha dos parâmetros foi feita utilizando uma análise exploratória dos clusters variando a distância mínima entre os registros ϵ e a densidade v . Para uma mesma quantidade de crimes, utilizando uma visualização da dispersão em um mapa, foram identificados os parâmetros ϵ , 500 metros e v , 3 vizinhos. Os mesmos valores foram encontrados por [Keler and Mazimpaka 2016] para um algoritmo similar ao DBScan, chamado OPTICS. Além disso, é feita a filtragem dos clusters, de acordo com uma quantidade mínima de crimes em cada cluster, para identificar as regiões mais afetadas pela criminalidade.

Após encontrar os pontos que pertencem a cada cluster, e remover os *outliers*, é necessário identificar os pontos das bordas de cada cluster. Como os clusters representam regiões, não é necessário considerar os pontos do interior das regiões. Para tal, é utilizado o método Convex Hull Graham [Kong et al. 1990], a indução do método se refere a encontrar um subconjunto mínimo de vértices de um polígono, sendo que, todos os vértices estejam dentro desse subconjunto. Nesse sentido, os vértices da borda do subconjunto estão dentro do polígono.

É viável utilizar a clusterização para identificar uma janela temporal capaz de compreender as mudanças da criminalidade durante o dia. Portanto, os parâmetros ϵ e v são fixados, e o tamanho da janela temporal variado de acordo com as horas, compreen-

endo: 1, 2, 4, 8, 12. Quão maior a janela temporal, maior a quantidade de crimes dentro da janela, e portanto, são identificadas regiões cada vez maiores. O tamanho dos clusters é de grande importância na identificação de regiões com alta incidência criminal pois, se um cluster compreender uma região vasta, não será possível identificar pontos críticos específicos dentro de uma cidade. Com a mesma visualização, em forma de mapa, utilizada para encontrar os parâmetros do algoritmo DBScan, foi possível reconhecer que a janela de tempo de 2 horas consegue identificar as mudanças na criminalidade das regiões.

Portanto, os clusters representam as áreas com alta incidência de acordo com um instante de tempo. Para um melhor entendimento do serviço proposto, a Figura 2 ilustra as mudanças na localização, distribuição dos clusters, como também no tamanho dos clusters durante uma rota. Em roxo está identificada a janela de tempo da rota a qual cada conjunto de clusters está relacionado. Assim sendo, em t_1 (Figura 2a) e t_2 (Figura 2b) é possível observar a existência de 3 clusters distintos, sendo que, os mesmos apresentam mudanças. Por fim, em t_3 (Figura 2c) é possível observar a ocorrência de 4 clusters.

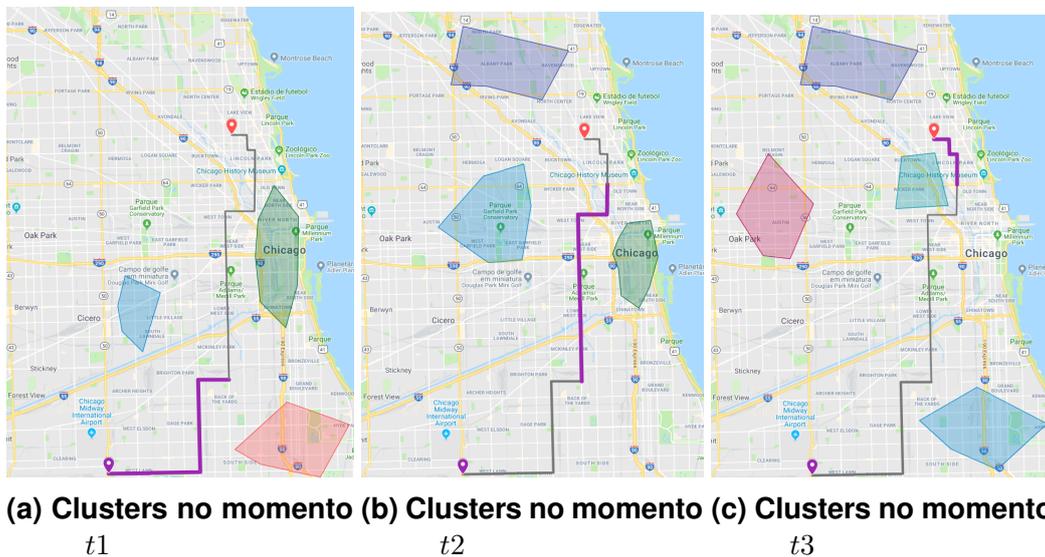


Figura 2. Variação da janela de tempo durante uma rota

4.2. Segurança Probabilística

A probabilidade pode ser descrita com uma função Gaussiana, sendo que, quão mais próximo do centroid de um cluster, menos segura é uma rota. Isso supondo que o centroid de cada cluster é o ponto com maior incidência criminal. Para considerar o tráfego da via, é possível utilizar o tempo de percurso já descrito durante a mineração das rotas. Dados os pontos presentes em uma rota, é possível verificar a distância para os clusters da janela temporal, e calcular a probabilidade do ponto estar dentro de cada cluster. Para calcular a segurança de cada ponto da rota, a solução proposta utiliza o *Parzen Window*. O benefício de utilizar esse método para o cálculo da probabilidade é a simples integração com uma *Kernel Function* Gaussiana, como descrita por [Babu and Viswanath 2008]. Com isso, é possível identificar o quão próxima uma rota está dentro de uma região insegura. Isto é, quanto mais próxima a rota está do cluster menos segura é tal rota. A Equação 1 descreve essa probabilidade, sendo:

$$\kappa = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x_i - x}{2\sigma^2}\right) \quad (1)$$

O parâmetro x representa a distância da borda até o centro do cluster, x_i a distância de cada ponto da rota até o centro do cluster, e σ o desvio padrão que caracteriza a suavização da distribuição Gaussiana. A suavização pode ser utilizada para caracterizar a densidade criminal em diferentes clusters. Por exemplo, um cluster com 100 crimes deve ter menor nível de segurança quando comparado com 10 crimes. Entretanto, ambos podem possuir o mesmo tamanho e formato, e ainda assim, densidades diferentes.

Para um melhor entendimento, a Figura 3 apresenta como a distância é utilizada no cálculo da probabilidade. A figura está dividida em passos para compreender cada fase do cálculo. Primeiramente é identificada a rota como os pontos em azul. Então, o polígono do cluster é representado em vermelho. O centro do cluster está representado na cor vinho, e são feitos os cálculos de duas distâncias. O valor de A representa a distância de um ponto da rota até o centro do cluster, e B representa a distância entre o ponto da rota até a borda do cluster. Portanto, na fórmula do *Parzen Window*, $x_i = A$, e $x = A + B$. Como o cluster é irregular, e não possui um formato circular, é necessário encontrar a borda mais próxima para caracterizar a proximidade com o centro do cluster.

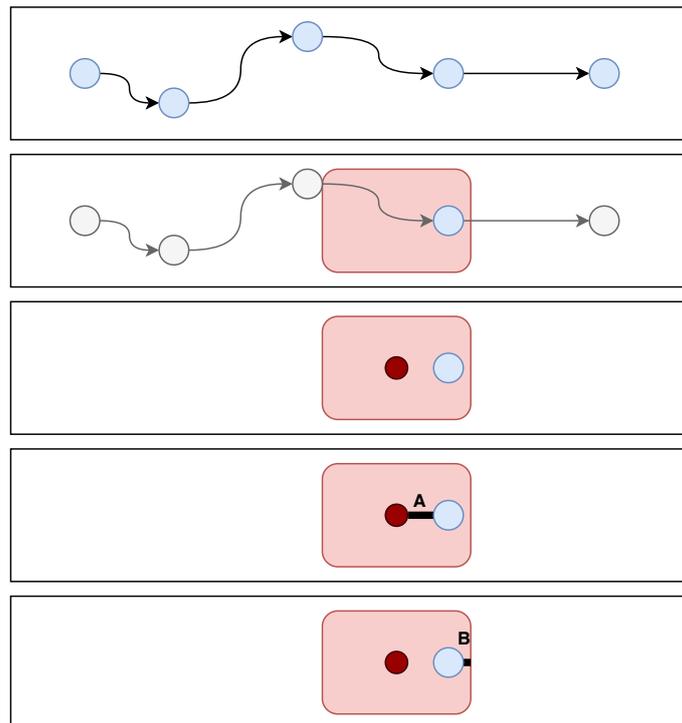


Figura 3. Cálculo da distância de cada ponto dentro do cluster

Como a segurança da rota é calculada em cada vértice, caso exista um cluster entre dois vértices, sendo que, nenhum deles pertence a esse cluster, ele não será considerado no cálculo da segurança. Para resolver esse problema é possível encontrar o ponto de cada aresta mais próximo de cada cluster, em seguida, verificar se esse ponto está dentro dos clusters. Para encontrar o ponto da aresta mais próximo de cada cluster é utilizada uma

interpolação entre o centro do cluster, em forma circular, e a aresta. Nessa interpolação, é desenhado um anel com raio igual a distância do centro até a aresta, e então, o ponto de intersecção é o ponto mais próximo. A segurança final da rota é caracterizada pela soma de todas as probabilidades como é descrito pela Equação 2.

$$\phi = \sum_{i=1}^n \kappa_i \quad (2)$$

Na equação, κ_i se refere a segurança de cada ponto, n é a quantidade de vértices e arestas, ou seja, $n = |V| + |E|$, e ϕ a segurança probabilística final da rota. Considerando a equação, quanto mais pontos dentro dos clusters mais insegura é a rota, ou seja, quão menor a probabilidade mais segura é a rota. No Algoritmo 1, pode ser observado um pseudocódigo do cálculo da segurança probabilística da rota. Como entrada é dada a rota (R) que deve ser calculada a segurança, a lista de todos os clusters identificados ($Clist$), e a hora do início da rota, para que seja possível identificar as mudanças na janela temporal. No algoritmo, $r \in R$ se refere a cada vértice e aresta de R . O valor de c representa todos os clusters de uma janela de tempo.

Algoritmo 1: CALCULA PROBABILIDADE

Entrada: $R, Clist, hora$
Saída: Segurança probabilística da rota

```

1 início
2   \\Inicialização das variáveis de estado
3    $prob = 0$ 
4    $tempo_{rota} = TEMPOINICIAL(R)$ 
5   para cada  $r \in R$  faça
6     para cada  $c \in Clist[hora]$  faça
7       \\Calcula a probabilidade e acumula na variável prob
8        $prob \leftarrow prob + PARZENWINDOW(r, c)$ 
9       \\Calcula o novo horário da rota baseado no tempo percorrido
          para aquele vértice
10       $tempo_{rota} \leftarrow tempo_{rota} + TEMPOPERCURSO(v'_i)$ 
11      \\Verifica se ocorreu a troca de janela temporal
12       $hora \leftarrow TROCAJANELA(tempo_{rota}, hora)$ 
13    fim
14  fim
15 fim
16 retorna  $prob$ 

```

5. Análise Experimental

A primeira base de dados se refere a utilização do módulo de Mineração de Rotas. A base de dados de rotas, foi construída utilizando os serviços da API do Here. A cada requisição é necessário passar um ponto geográfico de origem, um destino, se deve ser considerado o trânsito ou não, e a hora da rota a ser requisitada. O retorno de cada requisição se refere ao mapeamento das coordenadas de origem e destino para coordenadas reais nas vias, cada mudança de via feita pelo veículo, a distância e o tempo de percurso em cada via.

Para obter diferentes rotas ao longo do dia é necessário variar a origem e destino, como também a hora da requisição.

A base atual representa a requisição de 154 rotas à cada 15 minutos, durante o período de 7 dias seguidos. Portanto, foram mineradas 103488 rotas partindo de várias origens e destinos. Cada origem e destino representa, de forma geral, um distrito diferente da cidade de Chicago. Como a cidade de Chicago possui 77 distritos, então a cada 15 minutos é escolhida uma origem aleatória, e requisitadas rotas para todos os 77 distritos. Para facilitar a identificação de rotas distintas, são escolhidos aleatoriamente 2 pontos, em um mesmo distrito, e feitas as requisições. A cada conjunto de requisições de 1 ponto de origem para outros 77 de destino, são escolhidos novos pontos aleatórios dentro de cada distrito. A escolha aleatória da origem colabora com a fidelidade dos resultados, por não ser possível escolher rotas que possam favorecer os resultados do algoritmo.

Os dados criminais foram obtidos da iniciativa *Open Data Portal* da cidade de Chicago². Essa base de dados dispõe de uma ferramenta simples para filtrar, no próprio site, os anos e as informações referentes aos crimes. Por representar um ano completo, foram obtidos os dados do ano de 2017. A base de dados possui informações do tipo de crime, local (latitude e longitude), hora e data. O tipo de crime permite remover crimes que não sejam relacionados com o cenário de rotas veiculares. Alguns exemplos de crimes que são removidos pelo algoritmo são: *outros*, pela falta de identificação do crime, *fraude*, por não estar relacionado a ocorrência durante um percurso, e *obscenidade*, por ser um crime leve.

Com o intuito de analisar os resultados da utilização do serviço, foram obtidas todas as rotas e feita a correlação com os dados criminais agrupados pelo mês. Dentro da janela temporal de 2 horas, todas as rotas na janela foram mapeadas origem x destino, e a mais rápida, a mais curta, e a mais segura foram obtidas de cada origem x destino mapeada. As métricas de segurança utilizadas foram: a quantidade média de clusters atravessados (Figura 4), variação da probabilidade para cada tipo de rota (Figura 5a), o crescimento na probabilidade durante os passos das rotas (Figura 5b), e a probabilidade mínima, média, e máxima (Figura 6). A separação durante os meses do ano, permite verificar a variação da ocorrência de crimes durante o ano.

Na Figura 4, é possível observar a quantidade média de cluster atravessados por cada tipo de rota. Os dados criminais foram divididos entre os dias da semana e os dias do final de semana, sendo assim, é possível observar uma diminuição na quantidade de clusters identificados no final de semana. Como já é de se esperar, a rota mais segura atravessa menos clusters, sendo que, os valores menores que 1 se referem a possibilidade de encontrar rotas que não atravessam cluster algum durante o percurso todo. Durante os dias da semana, a maior diferença é possível observar no mês de agosto, com 1,35 cluster durante as rotas mais seguras, e um pouco mais de 2 clusters para as rotas mais curtas e mais rápidas. Nos dias do final de semana a média de clusters atravessados são, em todos os casos, menores que 0,5. No mês de março é possível observar a maior diferença, sendo de 0,20. Essas diferenças permitem identificar a capacidade de evitar clusters de forma a tornar um percurso mais seguro. Considerando a forma que é calculada a segurança das rotas, não é apenas verificada a capacidade de evitar clusters, mas a proximidade com o

²<https://data.cityofchicago.org/>

centro dos clusters é considerada.

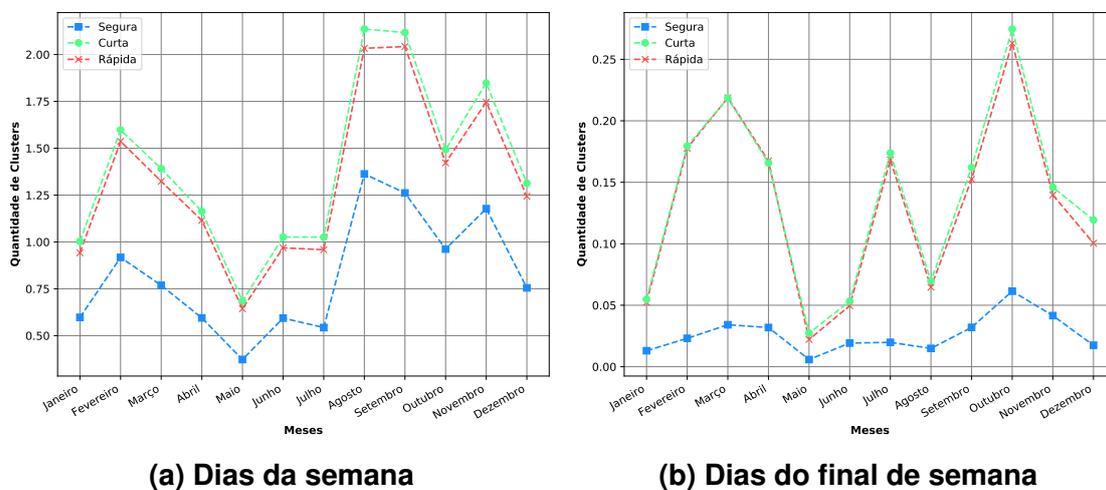


Figura 4. Quantidade média de clusters para os dias da semana e final da semana

Na Figura 5a, a diferença na segurança comparando as rotas mais seguras, as mais curtas, e as mais rápidas pode ser observada. O resultado obtido nas rotas mais seguras apresenta menor variação, comparada com os outros tipos de rotas, o que permite verificar uma melhor consistência na segurança. Além disso, a mediana para as rotas mais seguras é menor que o mínimo para os outros tipos de rotas. Ou seja, a diferença na segurança para os outros tipos de rotas é bastante considerável. Como a mediana não está no centro da variação interquartil, a maior parte das probabilidades é encontrada próxima ao limite inferior do interquartil, o que representa um resultado ainda melhor. Por fim, a diferença em termos de segurança, entre as rotas mais rápidas e as mais curtas, é pequena considerando a variação entre o máximo, o mínimo, e a mediana. Além disso, a variação interquartil de ambas possui quase o mesmo tamanho. Mesmo assim, as rotas mais rápidas apresentaram resultados melhores que as mais curtas.

Para verificar o crescimento da probabilidade durante as rotas, foram considerados cada ponto da rota (vértices e arestas) como passos. Então, foram obtidas 5 rotas aleatórias, de uma mesma janela de tempo, e verificadas as mudanças na probabilidade de cada rota. Na Figura 5b, são apresentados os resultados. A rota 2, que obteve no final o maior nível de insegurança, cruzou apenas 2 clusters, e possivelmente próximo ao centro pelo crescimento rápido na probabilidade. As rotas 1, 4, e 5 cruzam 3 clusters cada, mesmo assim não tiveram uma probabilidade final maior que a rota 2. Isso mostra que mesmo que atravessasse um maior número de regiões com alta incidência criminal, a distância para o centro dos clusters difere as rotas. Pela quantidade inicial de passos para obter o primeiro aumento na probabilidade, é possível identificar a proximidade com clusters, sendo que, até os 25 primeiros passos ocorreu aumento na insegurança. Após os 25 passos, as rotas mantiveram o mesmo nível de insegurança. Esse comportamento permite identificar a proximidade das origens para clusters naquela janela temporal. Mesmo sendo que não houve aumento na insegurança depois de 25 passos, podem ter outros clusters na janela temporal que não foram atravessados pelas rotas apresentadas.

Para analisar a variação da segurança probabilística, foi calculada a segurança de todas as rotas obtidas, sendo que, cada uma em sua janela temporal. Na Figura 6

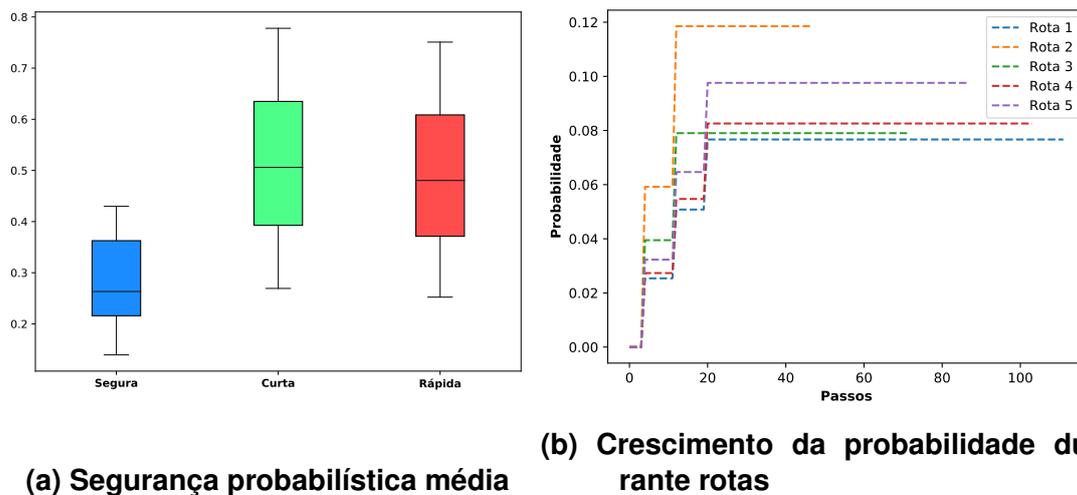


Figura 5. Métricas referentes a segurança probabilística

está apresentada essa distribuição das probabilidades encontradas para cada mês do ano. Como o Quartis tem a capacidade de reconhecer *outliers*, eles foram removidos dos dados. A média baixa das probabilidades se refere a como são encontrados os clusters, de forma a apenas identificar as regiões com alta incidência criminal. Portanto, é esperado que a maior parte das rotas esteja fora de regiões perigosas, por não serem consideradas todas as regiões que ocorreram crimes. Nos meses de Fevereiro e Abril é possível observar que a maior parte das rotas mineradas obtiveram probabilidade zero ou muito próxima de zero, pela diminuição na quantidade de crimes encontrada nesses meses. No mês de Agosto são obtidos os maiores níveis de insegurança, o que segue o que foi identificado na Figura 4a pela maior quantidade de clusters. A maior parte das probabilidades obtidas estão concentradas abaixo de 1,0, pela forma que é calculada a mesma. Como é calculada a média entre todos os pontos da rota, a cada ponto fora de qualquer clusters diminui a probabilidade final. Mesmo assim, é possível reconhecer a diferença probabilística entre as rotas e podem ser identificadas as mais seguras.

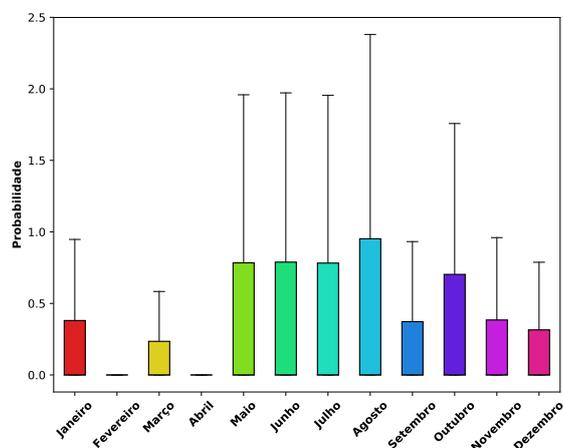


Figura 6. Probabilidades mensais

6. Conclusão

Nesse trabalho é apresentado um serviço para sugestão de rotas seguras, considerando dados criminais e rotas, de acordo com uma janela temporal. Esse serviço utiliza o algoritmo DBScan para identificar regiões com alta incidência criminal, são feitas requisições de rotas rápidas para considerar o tempo de percurso, e por fim, é utilizado *Parzen Window* para calcular a segurança probabilística de cada rota para sugerir a mais segura. Dependendo do caso, não é possível evitar 100% das regiões perigosas, por exemplo, se a origem ou o destino estão dentro de uma região perigosa, no entanto, o serviço permite a identificação desses casos de forma a minimizar o risco. Os resultados apresentados mostram essa capacidade do serviço, sendo que, uma menor quantidade de pontos deve estar dentro das regiões identificadas.

Agradecimentos. Os autores gostariam de agradecer o apoio financeiro concedido pelo CNPq (processo 167213/2018-0) para o bolsista de Mestrado Lucas Zanco Ladeira. Além desse, o apoio financeiro no projeto temático Big Cloud, processo nº 2015/24494-8 da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Referências

- Akabane, A. T., Gomes, R. L., Pazzi, R. W., Madeira, E. R. M., and Villas, L. A. (2017). Apolo: A mobility pattern analysis approach to improve urban mobility. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pages 1–6.
- ArcGIS (2018). Time windows. <http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/time-windows.htm>. Accessed: 2018-12-22.
- Babu, V. S. and Viswanath, P. (2008). An efficient and fast parzen-window density based clustering method for large data sets. In *2008 First International Conference on Emerging Trends in Engineering and Technology*, pages 531–536.
- Brennand, C. A. R. L., Boukerche, A., Meneguette, R., and Villas, L. A. (2017). A novel urban traffic management mechanism based on fog. In *2017 IEEE Symposium on Computers and Communications (ISCC)*, pages 377–382.
- Cunha, F., Maia, G., Celes, C., Santos, B., Rettore, P., Campolina, A., Guidoni, D., Souza, F., Ramos, H., Villas, L., Mini, R., and Loureiro, A. (2018). Sistemas de transporte inteligentes: Conceitos, aplicações e desafios.
- de Souza, A. M., Pedrosa, L. L. C., Botega, L. C., and Villas, L. (2018). Itssafe: An intelligent transportation system for improving safety and traffic efficiency. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–7.
- de Souza, A. M. and Villas, L. A. (2016). A fully-distributed traffic management system to improve the overall traffic efficiency. In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM '16*, pages 19–26, New York, NY, USA. ACM.
- de Souza, A. M., Yokoyama, R., Boukerche, A., Maia, G., Cerqueira, E., Loureiro, A. A., and Villas, L. A. (2016). Icarus: Improvement of traffic condition through an alerting and re-routing system. *Computer Networks*, 110:118 – 132.

- Farias, A. (2018). As roubadas enfrentadas por motoristas ao seguir atalhos de aparelhos de gps. <https://vejasp.abril.com.br/cidades/roubadas-motoristas-dicas-rota-gps-waze/>. Accessed: 2018-11-21.
- Galbrun, E., Pelechrinis, K., and Terzi, E. (2016). Urban navigation beyond shortest route. *Inf. Syst.*, 57(C):160–171.
- Johnson, I., Henderson, J., Perry, C., Schöning, J., and Hecht, B. (2017). Beautiful…but at what cost?: An examination of externalities in geographic vehicle routing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(2):15:1–15:21.
- Keler, A. and Mazimpaka, J. D. (2016). Safety-aware routing for motorised tourists based on open data and vgi. *Journal of Location Based Services*, 10(1):64–77.
- Kim, J., Cha, M., and Sandholm, T. (2014). Socroutes: safe routes based on tweet sentiments. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*.
- Kong, X., Everett, H., and Toussaint, G. (1990). The graham scan triangulates simple polygons. *Pattern Recognition Letters*, 11(11):713 – 716.
- Pan, J., Popa, I. S., and Borcea, C. (2017). Divert: A distributed vehicular traffic re-routing system for congestion avoidance. *IEEE Transactions on Mobile Computing*, 16(1):58–72.
- Quercia, D., Schifanella, R., Aiello, L., and Mclean, K. (2015). Smelly maps: The digital life of urban smellscapes. *Proceedings of 9th International AAAI Conference on Web and Social Media (ICWSM2015)*.
- Quercia, D., Schifanella, R., and Aiello, L. M. (2014). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 116–125, New York, NY, USA. ACM.
- Santos, F. A., Rodrigues, D. O., Silva, T. H., Loureiro, A. A. F., Pazzi, R. W., and Villas, L. A. (2018a). Context-aware vehicle route recommendation platform: Exploring open and crowdsourced data. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–7.
- Santos, F. A., Silva, T. H., Loureiro, A. A. F., Boukerche, A., and Villas, L. A. (2018b). Identificação da reputação de Áreas urbanas externas com dados de mídias sociais. *Anais do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, 36.
- Shah, S., Bao, F., Lu, C.-T., and Chen, I.-R. (2011). Crowdsafe: Crowd sourcing of crime incidents and safe routing on mobile devices. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 521–524, New York, NY, USA. ACM.
- Taha, A. M. (2017). Facilitating safe vehicle routing in smart cities. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–5.
- Utamima, A. and Djunaidy, A. (2017). Be-safe travel, a web-based geographic application to explore safe-route in an area. *AIP Conference Proceedings*, 1867(1):020023.