

# Avaliação do uso de um modelo de criminalidade na determinação de rotas seguras para ônibus urbanos

Vinícius G. J. Almeida<sup>1</sup>, Thais R. M. B. Silva (Orientadora)<sup>1</sup>,  
Fabrício A. Silva (Coorientador)<sup>1</sup>

<sup>1</sup> Instituto de Ciências Exatas e Tecnológicas  
Universidade Federal de Viçosa (UFV) – Campus Florestal  
Florestal, MG – Brasil

{vinicius.jesus, thais.braga, fabricio.asilva}@ufv.br

**Resumo.** *O trânsito ciente de contexto é uma linha de pesquisa que, nos últimos anos, ganhou relevância devido ao aumento de dispositivos e aplicações IoT relacionadas. Grande parte dos trabalhos nessa área propõem abordagens para veículos particulares, deixando de lado investigações ligadas ao transporte público. Neste trabalho foi implementada e avaliada uma estratégia genérica de construção de rotas de ônibus mais seguras. A solução construída, além de gerar trajetos mantendo os pontos de parada originais, também propõe melhorias realocando alguns destes para regiões mais seguras. Os resultados para os pontos de paradas realocados mostraram ser possível reduzir em até 20% o tráfego rodoviário em regiões inseguras, com aumento do trajeto em apenas 13%.*

## 1. Introdução

Devido aos recentes avanços tecnológicos nas áreas de Internet das Coisas (IoT) e sistemas distribuídos [Atzori et al. 2018], dispositivos capazes de coletar dados de forma prática e eficiente passaram a fazer parte da rotina das pessoas. Com isso, novas oportunidades de negócios e estudos surgiram, possibilitando a criação de modelos de dados a partir de algoritmos de aprendizado de máquina que visam inferir, obter conhecimento e auxiliar em tomadas de decisão em muitas áreas [Verbraeken et al. 2020]. Dentre essas, as aplicações para cidades inteligentes se apresentam como promissoras, oferecendo diferentes oportunidades, tais como aquelas ligadas ao trânsito ciente de contexto [Zhang et al. 2011] que, de modo geral, se refere ao uso de dados para oferecer melhorias e comodidades ao trânsito, tanto para motoristas quanto para o fluxo de veículos.

Em um futuro próximo serão os próprios dispositivos computacionais ligados às aplicações que farão, de forma autônoma, a coleta e uso de dados contextuais, aplicando-os na personalização e tomada de decisões [Atzori et al. 2016]. Em particular, Sistemas de Transporte Inteligentes (STI) utilizarão um ou mais modelos de aprendizado de máquina para traçar rotas a serem utilizadas pelos mais diferentes tipos de veículos [Ladeira et al. 2019]. Vários trabalhos na literatura associam dados de diversas circunstâncias ao trânsito, como: riscos de acidentes nas vias, condições climáticas e níveis de periculosidade ao trafegar. Parte deles envolvem esse último citado e utilizam modelos de criminalidade como um dos tipos de contextos a serem considerados, Seção 2. Estes modelos contribuem para identificar vias e regiões com alto índice de registros criminais, possibilitando recomendações de rotas que evitam regiões inseguras de uma cidade.

Embora os trabalhos sobre rotas seguras encontrados até o momento tenham avaliado diferentes soluções, as mesmas foram testadas majoritariamente para veículos particulares e avaliadas sob o ponto de vista computacional, ou seja, com o foco em criar soluções eficientes que evitem ao máximo regiões perigosas, e de fácil expansão e reuso. Entretanto, no contexto de transportes públicos, estas soluções podem abrir mão de questões como o impacto social causado, neste caso, deixando pessoas sem acesso ao meio de transporte. Diante disso, este trabalho difere dos demais por avaliar alternativas para veículos de transportes públicos, em particular, ônibus municipais. Desta forma, por se tratar de um recorte que envolve diversas pessoas e questões sociais, algumas limitações quanto às recomendações devem ser avaliadas, como a área de cobertura das rotas de ônibus, por exemplo. Afinal, dependendo das alternativas apresentadas, regiões da cidade podem ser prejudicadas e ficar sem o acesso ao meio de transporte.

Diante do que foi apresentado, tem-se para este trabalho a seguinte pergunta: é possível que o sistema rodoviário de uma grande cidade utilize soluções de rotas seguras equilibrando os ganhos de nível de segurança dos passageiros com o aumento da distância percorrida pelas rotas? Além disso, com uma possível implementação de cálculo de rotas seguras que evitam áreas consideradas perigosas, o sistema ainda seria capaz de manter-se comprometido com a área de cobertura de transporte? Em outras palavras, algumas pessoas correriam o risco de ficar sem acesso ao transporte em algumas regiões? A hipótese inicial é de que seja possível, desde que sejam observados limites acerca do que possa ser modificado em termos de pontos de paradas e as trajetórias entre eles.

O objetivo deste trabalho é avaliar uma das soluções específicas de elaboração de rotas seguras já propostas na literatura [Santos et al. 2017] e [Ladeira et al. 2019] para o sistema rodoviário da cidade de São Paulo. Para isso, o foco é avaliar o impacto da mesma sobre métricas que indicam nível de segurança dos passageiros, distância percorrida nas rotas e manutenção das áreas de cobertura, bem como a relação entre as mesmas.

O restante deste trabalho está dividido da seguinte forma: a Seção 2 possui uma breve descrição dos principais trabalhos relacionados encontrados na literatura; Detalhes sobre o modelo de determinação de rotas seguras da literatura utilizado neste trabalho para rotas rodoviárias, bem como a ampliação proposta ao mesmo, podem ser encontrados na Seção 3. Os resultados que mostram a relação entre segurança e distância percorrida pelos ônibus com e sem o uso dos modelos de rotas seguras estão na Seção 4. Por fim, a Seção 5 contém comentários finais e alguns direcionamentos para trabalhos futuros.

## **2. Trabalhos Relacionados**

Diversos trabalhos na literatura têm, nos últimos anos, apresentado soluções que visam oferecer rotas calculadas com base em aspectos que ultrapassam apenas o uso da distância, tais como índices de acidentes e registros de criminalidade.

Considerando rotas para pedestres, o trabalho de [Levy et al. 2020] apresenta SafeRoute, uma solução baseada no uso de aprendizado profundo por reforço para auxiliar pessoas a se locomoverem por regiões mais seguras, evitando áreas com elevado número de registros de assalto e assédio. O modelo, treinado com informações advindas de recentes bases de dados de criminalidade, foi testado em 3 diferentes grandes cidades. Na comparação com outros métodos do estado da arte, SafeRoute obteve resultados que mostram ser possível aumentar a distância entre pessoas e locais de crimes e ainda

oferecer trajetórias mais curtas entre origem e destino. Já SafeRNet [Liu et al. 2017] é um *framework* baseado no uso de redes bayesianas que visa diminuir ocorrências de acidentes para veículos particulares. No estudo foram utilizados critérios como condições climáticas, fluxo de veículos, colisões e condições das vias para propor variações de rotas ao decorrer do dia.

Em particular, existem muitos estudos que apresentam propostas para confecção de rotas para veículos particulares, considerando especialmente o índice de segurança das mesmas. [Félix Mata 2016] apresentam um *framework* que recomenda rotas seguras com base em um sistema de recomendação híbrido que utiliza dados oficiais e de *tweets*. Este é construído através de redes bayesianas para auxiliar motoristas a trafegarem por regiões com altos índices e históricos de registros criminais. O modelo foi treinado com uso de duas fontes de dados (criminais públicos e *tweets*) visando estabelecer correlações. O trabalho obteve precisão de cerca de 75% nas recomendações de rotas seguras utilizando apenas os dados da rede social. Além disso, também apresenta flexibilidade quanto à janelas temporais. SafePaths [Galbrun et al. 2016] é um sistema de recomendação de rotas seguras que utiliza clusterização *Gaussian Kernel Density Estimation (KDE)*. As rotas são geradas a partir de pequenas regiões como ruas e bairros, tendo sido utilizadas como base as cidades de Chicago e Philadelphia, ambas com *datasets* criminais atualizados e repletos de informações relevantes. Os autores argumentam que os resultados mostraram que a solução é robusta em relação a tempo de execução e taxa de atualização dos dados criminais. CERVA [Ladeira et al. 2020] é também uma solução para oferecer alternativas de rotas e filtros baseada em clusterização KDE e grafos ponderados. A proposta apresentada é baseada na intensidade do tráfego da via juntamente com sua condição de segurança para diferentes janelas temporais do dia. O trabalho conta com bases de dados atualizadas, referentes à cidade de Chicago, incluindo informações criminais oficiais e condições de tráfego provenientes de APIs de sistemas reais, como o *Here Maps*.

Por fim, o trabalho de [Tompson et al. 2009] apresenta *Hot Routes*, uma proposta voltada para o tráfego rodoviário. A solução, baseada em clusterização KDE, tem como principal objetivo identificar regiões inseguras, referentes à criminalidade, para rotas de ônibus, como pontos de paradas ou trechos do percurso, na cidade de Londres e exibi-las em um mapa.

Como pôde ser observado, a grande maioria dos trabalhos mencionados utilizam estratégias para contextos específicos, no que diz respeito à localidades e tipo de dados considerados. Além disso, boa parte das propostas está relacionada a veículos particulares. Dentre as soluções apresentadas acima, aquelas que mais se assemelham ao trabalho proposto neste artigo são as de identificação de caminhos inseguros com base em métodos de clusterização. Destes, [Tompson et al. 2009] é o único relacionado à linhas de transportes coletivos. Entretanto, diferente do presente trabalho, não propõe novas sugestões de rotas para substituir as identificadas como inseguras. Além disso, também não possui nenhum outro critério de remoção de pontos de paradas ou trechos da rota além da periculosidade.

### 3. Utilizando um Modelo de Criminalidade na Determinação de Rotas Seguras para Ônibus Urbanos

Este trabalho consiste na implementação de uma solução da literatura [Ladeira et al. 2019] que visa a identificação de rotas seguras para veículos particulares no contexto de transportes rodoviários. Em particular, é utilizada a subcategoria de ônibus urbanos públicos de uma grande metrópole, qual seja, a cidade de São Paulo. Por se tratar de uma aplicação para transporte coletivo, algumas características da solução base precisaram ser adaptadas, como, por exemplo, a necessidade de manter os locais de embarque e desembarque de passageiros (pontos de paradas) dentro das rotas calculadas. Com isso, as escolhas para a construção de rotas seguras precisaram ser feitas entre os pontos de paradas e não apenas considerando os pontos de origem e destino dos veículos, para esta adaptação foi necessário apenas conectar os trechos entre pares de pontos de ônibus. Comparando as rotas seguras geradas à outras formas alternativas de cálculo das mesmas foi possível identificar impactos em termos de índices de segurança, distância percorrida e área de cobertura. Com o objetivo de analisar o impacto de uma implementação ainda mais ampla da ideia proposta pelo trabalho da literatura no contexto dos ônibus urbanos [Tompson et al. 2009], foi proposta uma modificação na solução em que pontos de paradas pudessem ser realocados. Porém, como este trabalho também considera o impacto social como parte da solução, são levados em consideração algumas condições, detalhadas na Subseção 3.3, que visam não prejudicar a população local da região analisada antes de se propor mudanças nas configurações originais das linhas.

#### 3.1. Roteamento Seguro de Veículos Particulares

Como *baseline* foi utilizado o trabalho [Ladeira et al. 2019]. Este é uma continuação de outros trabalhos, tais como [Santos et al. 2017], oriundos de um mesmo grupo de pesquisa. [Ladeira et al. 2019] propuseram um sistema de recomendação de rotas seguras baseado em dados históricos criminais. O trabalho propõe o uso do algoritmo DBScan para identificar áreas de alta criminalidade e um algoritmo que sugere rotas, em janelas temporais, que evitam passar pelas mesmas. Foram utilizados conjuntos de dados da cidade de Chicago e apresentados resultados que mostram que a mediana do índice de criminalidade para as rotas mais seguras é menor que o mínimo das demais comparadas (mais curtas e mais rápidas).

Dentre as diferentes opções disponíveis de algoritmo de aprendizagem de máquina para agrupamento de dados, os autores do *baseline* optaram pelo algoritmo *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) [DBSCAN] para a identificação de áreas criminais geograficamente distribuídas pela cidade. Esta escolha foi justificada devido à habilidade do algoritmo em gerar agrupamentos (clusters) com diferentes formatos e tamanhos e por ter bom resultado de clusterização mesmo na presença de ruídos [Santos et al. 2017]. Por se tratar de um algoritmo que mantém os nós originais nas suas respectivas posições, o processo de clusterização precisa apenas relacionar estes nós. Isso é feito através de dois parâmetros. O primeiro determina o raio máximo que um nó vizinho deve estar para fazer parte de um mesmo cluster. Este valor é representado por  $\epsilon$ . Já o segundo, identificado como  $\nu$ , determina a quantidade mínima de nós vizinhos, contidos em um único agrupamento, necessária para que um *cluster* seja classificado como válido. Para este trabalho os autores utilizaram uma base de dados criminal

da cidade de Chicago para o ano de 2017, que contém crimes de diversas naturezas, totalizando 268.793 ocorrências criminais. Dadas essas informações, definiram  $\varepsilon = 500\text{m}$  e  $\nu = 3$ . Estes valores foram obtidos após uma análise exploratória variando os dois parâmetros principais do algoritmo [Ladeira et al. 2019].

Para relacionar as informações criminais e avaliar as rotas, os autores utilizaram a estrutura de dados grafo para representar a cidade escolhida. Para fazer este mapeamento foi utilizado um grafo dirigido com os vértices representando as interseções (esquinas) e as arestas representando as vias da cidade.

A estratégia utilizada para seleção de rotas a serem avaliadas foi utilizar aquelas consideradas mais rápidas disponibilizadas pelo Here API [Here]. Ao todo, foram coletadas 103.488 rotas para diferentes origens e destinos da cidade. Com isso, não é necessário criar rotas sintéticas (artificiais) para validar os modelos criados. Desta forma, diante dessas rotas foram aplicadas as métricas para identificação de rotas seguras.

Para o cálculo das rotas seguras os autores do *baseline* elaboraram uma equação (Equação 1) que leva em consideração o resultado do agrupamento dos dados criminais realizado pelo algoritmo DBScan. Esta equação é composta por um somatório que classifica a distância de cada nó da rota em relação ao centro de *clusters* mais próximo. Com isso, ao final da iteração, o valor retornado  $k$ , recebe uma pontuação que representa a periculosidade da rota. Desta forma, quanto menor for este valor, mais segura será a rota. Para se obter este resultado, essa equação denominada como *Parzen Window*, é uma forma não paramétrica para estimar a função densidade de probabilidade de uma variável aleatória. O benefício de utilizá-la é pela simples integração com uma *Kernel Function Gaussiana*, que representa uma função de densidade de probabilidade caracterizada pela forma simétrica de curva de sino que descreve distribuições normais [Babu and Viswanath 2008].

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

A Equação 1 conta com os seguintes parâmetros:  $k$ , que representa o resultado do somatório. Este valor acumula a pontuação de segurança de cada nó da rota. Desta forma, se torna um parâmetro de fácil comparação e interpretação, cujo significado é: quanto maior, menos segura é uma determinada rota. Logo, através dele temos a pontuação geral da rota;  $n$  indica a quantidade de vértices (representados por esquinas) de uma determinada rota entre o ponto de partida e a chegada;  $\sigma$  representa o desvio padrão da quantidade de crimes por *cluster* criminal;  $x_i$  representa a distância de cada vértice da rota em relação ao centro do *cluster* mais próximo; e  $x$  representa a distância da borda do *cluster* mais próximo em relação ao seu centro.

### 3.2. Aplicação do *Baseline* no Cálculo de Rotas Seguras para Ônibus Urbanos

A cidade de São Paulo foi escolhida neste trabalho uma vez que foi a única que possuía três requisitos essenciais ao desenvolvimento do mesmo, que são: a disponibilidade da cidade escolhida no grafo da biblioteca utilizada para traçar rotas, esta será detalhada a seguir; Uma base de dados criminal que seja atualizada, com geolocalização e breve

descrição do crime <sup>1</sup>; E detalhes de rotas de transporte público, como informações das linhas rodoviárias e seus pontos de paradas <sup>2</sup>. Os dois últimos tópicos se mostraram mais complicados, pois restringiram a escolha da cidade para apenas São Paulo - SP, única, até onde pudemos pesquisar, que possuía todos os três requisitos bem definidos e detalhados.

Ao replicar os passos dos artigos de *baseline* foram identificados alguns problemas, sendo necessário reajustar alguns parâmetros do algoritmo para corrigi-lo e contextualizá-lo ao cenário deste trabalho. Esta adaptação foi necessária pois se fossem consideradas apenas uma ou poucas ocorrências em cada região, os clusters resultantes preencheriam todas as regiões da cidade, chegando a conclusão de que não há nenhum local seguro para trafegar. Contudo, para contornar este problema foram estabelecidos alguns critérios, pois, se tratando de clusters formados através do algoritmo DBSCAN, um número mínimo de vizinhos necessários deve ser definido. Sendo assim, como a base de dados criminal possui crimes registrados no decorrer de um ano, uma alternativa é estabelecer um número mínimo de ocorrências mensais, coerente com o significado de uma área ser segura ou insegura, para então aplicar para todo o ano. Com isso, foi estabelecido um valor que representa uma média de pelo menos três crimes a cada mês, o que resulta em pelo menos 36 crimes anuais em cada região avaliada,  $v = 36$ . Desta forma, todo agrupamento que conter pelo menos 36 ocorrências criminais será considerado uma região insegura. Além disso, o outro parâmetro relacionado à distância máxima entre dois nós para fazerem parte de um mesmo *cluster* foi definido para 100 metros,  $\epsilon = 100m$ .

Após o ajuste comentado anteriormente, já foi possível prosseguir na implementação do trabalho. A seguir serão discutidas e apresentadas todas as etapas da replicação do *baseline* junto aos ajustes para o contexto de linhas de ônibus.

O primeiro passo na construção da solução proposta neste trabalho é o uso do “OSMnx: Python for street networks” [OSMnx], que é uma ferramenta utilizada na construção do grafo que representa a cidade de São Paulo. Esta ferramenta está disponível como um pacote da linguagem Python. O grafo da cidade de São Paulo - SP se encontra disponível e pronto para uso. Nele, cada esquina da cidade é representada como um vértice, sendo as ligações entre elas modeladas como arestas direcionais.

O *dataset* criminal é um arquivo no formato CSV (Comma-Separated Values) obtido através da plataforma *data.world* [DataWorld], contendo aproximadamente 770 mil registros criminais da cidade de São Paulo no ano de 2016. Este arquivo engloba colunas como: número do boletim de ocorrência, data do crime, localização descrita e através de latitude e longitude, rubrica do crime, entre outros. Com base nessas informações, foi feito um primeiro ajuste relacionado à localização dos crimes, ou seja, por se tratar de um sistema que depende de clusterização, dados sem geolocalização não são úteis para o trabalho, podendo ser descartados. Com isso, após essa remoção, a base de dados criminais foi reduzida para aproximadamente 728 mil crimes. Outro ponto importante está relacionado às características dos crimes com o contexto do problema, ou seja, por se tratar de um estudo relacionado ao trânsito, alguns tipos criminais não interferem diretamente na segurança de rotas veiculares, como por exemplo: “Oferecer droga a pessoa de seu relacionamento (Art.33,§3º)”. Logo, estes tipos de crimes foram desconsiderados. Dentre essa análise de crimes relacionados, foram mantidos 4 tipos criminais dos 36 existentes

---

<sup>1</sup>Dataset criminal: <https://data.world/maszanichi/boletins-de-ocorrencia-sp-2016>

<sup>2</sup>Dataset das rotas de ônibus: [https://interscity.org/open\\_data/](https://interscity.org/open_data/)

na base de dados, os quais podem ser observados na Tabela 1.

<b>Tipos de crimes</b>	<b>Quantidade</b>	<b>Porcentagem (%)</b>
Total de crimes válidos	728305	100
Total de crimes selecionados	587018	80,60
<b>Detalhes dos tipos de crimes selecionados</b>	<b>587018</b>	<b>100</b>
Furto (art. 155)	278971	47,52
Homicídio simples (art. 121)	2913	0,50
Roubo (art. 157)	305106	51,98
Tráfico de entorpecente (Art. 12)	28	0,005

**Tabela 1. Distribuição dos tipos de crimes selecionados e utilizados**

Com isso, após esta etapa de seleção de crimes coerentes para o contexto analisado, basta utilizá-los em conjunto com o algoritmo DBScan e gerar e distribuir os *clusters* criminais pela cidade. A Figura 1a exemplifica o que foi mencionado acima. Observe que os clusters não possuem formatos regulares, pois seguem estritamente as localizações reais dos crimes analisados. Já a Figura 1b é uma versão ampliada de um trecho do grafo para permitir uma visualização mais detalhada de alguns *clusters*.



**(a)** Mapa completo.



**(b)** Trecho da cidade.

**Figura 1. Grafo da cidade de SP com *clusters* criminais.**

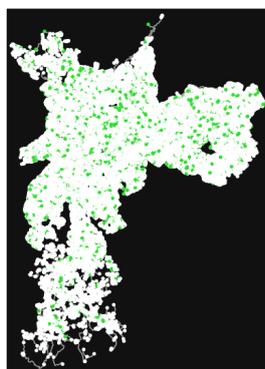
Um segundo conjunto de dados utilizado contém as informações sobre as rotas de ônibus da cidade de São Paulo - SP. Estes dados foram retirados da plataforma *Interscity* [InterSCity]. Dentre as informações contidas, estão presentes dados como: linhas de ônibus para cada dia da semana e também para feriados, além de possuir os pontos de ônibus distribuídos pela cidade com suas respectivas geolocalizações. Estas informações estão distribuídas através de dois modelos de arquivos XML. O primeiro é separado de acordo com os dias da semana. Dentro destes arquivos, existem elementos com atributos como o identificador das linhas de ônibus, que totalizam 2183 linhas, com suas respectivas listas de pontos de parada. Já o segundo arquivo é um complemento do primeiro, pois nele estão armazenados os pontos de paradas com suas respectivas coordenadas geográficas,

totalizando 20.041 pontos de parada. A Tabela 2 apresenta estatísticas da distribuição de pontos de paradas por linhas de ônibus. Em média, percebe-se que cada linha possui aproximadamente 43 pontos de parada.

Mínimo	Mediana	Máximo	Média	Desvio padrão
5	40	132	42,85	19,90

**Tabela 2. Estatísticas de pontos de parada da base de dados de linhas de ônibus.**

A primeira tentativa de inserir os pontos de ônibus no grafo visava inserir novos nós e arestas no grafo, caso fosse necessário, pois cada ponto possui sua respectiva localização exata. Entretanto, na tentativa de implementar esta abordagem foi identificado que a biblioteca não possui suporte para este tipo de operação. Desta forma, a solução aproximada escolhida foi identificar quais nós (esquinas) estavam mais próximos dos pontos de ônibus originais e classificá-los como pontos de paradas. Foi estabelecido um padrão de cores para diferenciá-los de nós comuns. Os vértices de cor verde representam nós com pontos de parada, já os de cor branca, representam esquinas normais. O resultado desta adaptação pode ser observado nas Figuras 2a e 2b.



**(a)** Mapa completo.



**(b)** Trecho da cidade.

**Figura 2. Grafo da cidade de SP com pontos de paradas de ônibus.**

Vale a pena ressaltar que uma funcionalidade para o cálculo de rotas está disponível em um módulo da biblioteca OSMnx, sendo que para utilizá-la foi necessário apenas mapear os identificadores e geolocalizações dos pontos originais com as novas referências criadas no grafo.

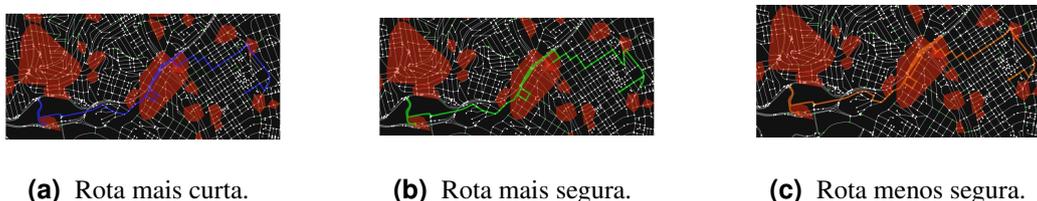
O processo de identificação de rotas seguras foi construído separando a rota completa (entre pontos de origem e destino de uma linha de ônibus) em rotas entre pares de pontos de parada. O principal motivo deste desmembramento está relacionado com o uso da Equação 1, que pondera a quantidade de vértices percorridos como critério de segurança. Desta forma, dividir a rota completa em trechos menores faz com que haja um detalhamento maior para cada par de pontos de parada consecutivos, além de penalizar

também o número de paradas para embarque e desembarque de passageiros, que também pode ser considerada como fator de segurança, visto que o ônibus fica mais vulnerável à crimes quando está parado.

Com estes critérios adotados, a estratégia utilizada foi usar a própria ferramenta OSMnx [OSMnx] para gerar diferentes opções de rotas, sempre entre cada dois pontos de parada consecutivos, para cada linha de ônibus. Para cada rota, cada um de seus vértices são aplicados à Equação 1, dependendo apenas de identificar o cluster criminal mais próximo. Desta forma, ao final do processo uma lista de rotas ranqueadas é gerada para cada trecho da rota completa, classificando cada opção de caminho com uma pontuação  $K$ . Com isso, basta somar todos estes valores para obter a pontuação da rota completa, para então compará-la às demais.

Para a avaliação do uso do *baseline* no contexto das linhas de ônibus, foi necessário traçar 3 tipos de rotas para cada linha: rotas mais seguras, rotas menos seguras e rotas mais curtas. A rota mais curta é utilizada como um parâmetro de comparação para as demais rotas geradas, pois, através dela é possível criar comparações de distâncias percorridas e avaliar se os resultados são coerentes. Sendo assim, para construí-la, o único fator analisado é a menor distância percorrida entre um ponto de parada e outro, desconsiderando o fator criminal. Contudo, como esta rota é gerada da mesma forma que as demais, apenas com a ressalva de menor distância possível, identificar o grau de segurança é idêntico ao das outras rotas. Logo, também torna-se possível classificá-la através de um *ranking*. Um exemplo de rota mais curta pode ser observado na Figura 3a. Esta rota é a rota de ID “407T-10-0” do *dataset*, a qual foi escolhida aleatoriamente apenas para ser utilizada nas figuras de exemplo a seguir.

Diferente da rota mais segura, os demais dois tipos de rotas restantes são construídas apenas analisado o  $K$  obtido para cada trecho. Desta forma, para cada par de pontos de parada são escolhidas, dentre a lista de opções, a rota com menor  $K$  para a rota mais segura e a rota com maior  $K$  para rota menos segura. Logo, utilizando este conceito é possível contemplar os dois extremos de segurança entre cada opção de linha de ônibus. Um exemplo de resultado utilizando a mesma linha de ônibus da Figura 3 pode ser observado nas Figuras 3b e 3c.



**Figura 3. Rotas de ônibus para a linha de ID “407T-10-0”.**

### 3.3. Ampliação do Modelo de Rotas Seguras

Diante da solução apresentada anteriormente, seguindo o *baseline* [Ladeira et al. 2019], as opções de rotas criadas são geradas com base nos pontos de parada originais, alterando apenas os caminhos entre uma parada e outra consecutiva. Entretanto, ao construir este modelo foi observado um fator prejudicial para a identificação de rotas mais seguras.

Foi observado que independente das possibilidades de rotas criadas, em alguns contextos, determinados pontos de paradas seriam os locais menos seguros, pois estariam localizados dentro de um *cluster* criminal.

Com isso, foi adotada uma solução que visa estender a solução inicial, com o intuito de criar novas alternativas mais seguras baseadas no cenário corrente da cidade de São Paulo. A estratégia utilizada para mitigar o problema apresentado é tentar remover um ponto de ônibus localizado dentro de um *cluster* criminal e utilizar outro, próximo, para substituí-lo. Para isso, foram adotados alguns critérios.

Para determinar os critérios, o principal tópico analisado foi o fator social, pois o trabalho visa não prejudicar a população que reside próxima a lugares com regiões perigosas. Para que um ponto de ônibus candidato, localizado em um *cluster*, seja removido, deve haver algum outro ponto fora de uma região criminal e localizado dentro de um raio  $R$  máximo. No trabalho atual,  $R = 250m$ . Além disso, por se tratar de transporte terrestre, também foi considerado a distância real de locomoção, isto é, a distância percorrida de uma origem para um destino passando pelas vias existentes entre eles na cidade. Desta forma, esta distância foi definida como 1000 metros, do ponto original até o ponto candidato, visto que o deslocamento real possui uma distância maior do que aquela em linha reta. Para aplicar esta estratégia de percurso foi considerado o fator de acesso rodoviário, ou seja, é preciso que uma pessoa consiga ir e voltar ao ponto candidato partindo do ponto original, pois, com isso é possível garantir que as pessoas não terão que percorrer distâncias maiores que as estabelecidas para conseguir o acesso ao transporte.

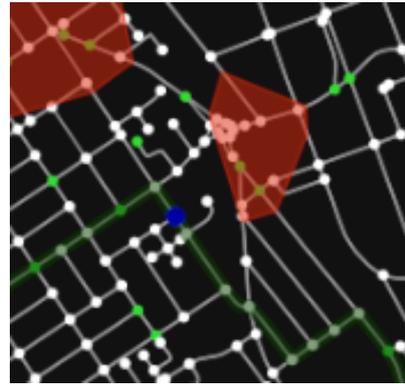
A nova proposta de implementação de rotas segue os critérios detalhados anteriormente apenas para propor novos pontos de parada, caso seja possível. Se algum dos critérios não forem estabelecidos ou se o ponto de parada não estiver localizado em nenhuma região criminal, o mesmo continuará sendo o ponto original. Após identificar e substituir os pontos originais localizados em *clusters* criminais por candidatos válidos, a implementação para identificação de rotas é idêntica ao modelo original.

Para fins de comparação a Figura 4a contém um trecho de uma rota com o ponto original, destacado pela cor azul, localizado dentro de um *cluster*. De acordo com o que foi mencionado anteriormente, este ponto pode ser analisado para receber um substituto localizado em uma região segura. A estratégia classifica possíveis candidatos de acordo a distância que estão do ponto original, tendo como limite os valores já apresentados. Com isso, ao analisar a Figura 4a é possível perceber que existe um ponto de ônibus adjacente ao ponto original, entretanto, este também se encontra dentro de um *cluster*, logo, é descartado. Com isso, outros pontos são avaliados e o que possui a menor distância em relação ao original e está em uma região segura é o ponto azul representado na Figura 4b. A partir desta estratégia uma nova rota é calculada para este trecho, substituindo a calculada a partir do ponto de parada original.

Após aplicar todas as etapas mencionadas, novas rotas são geradas considerando os novos pontos de parada, caso seja possível e necessário. Assim como na solução anterior, também são considerados as três formas de se traçar rotas mais seguras, menos seguras e mais curtas. A rota calculada no exemplo anterior é uma aplicação do algoritmo para identificar a rota mais segura considerando as possíveis alterações dos pontos originais. A rota completa pode ser observada na Figura 5.

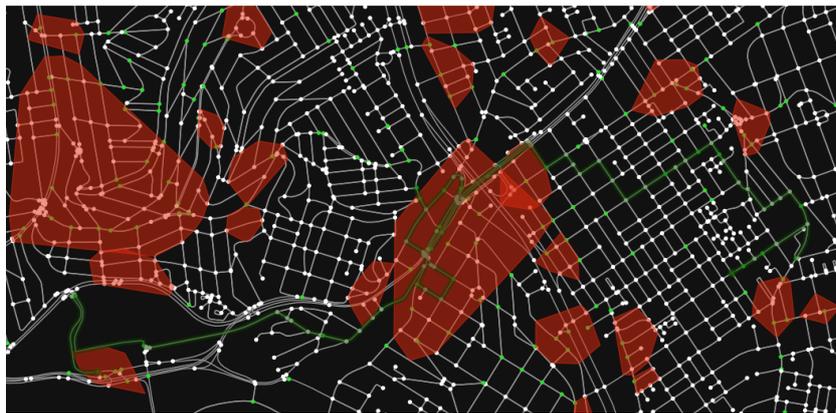


(a) Parada original dentro de *cluster* criminal.



(b) Parada realocada para uma região segura.

**Figura 4. Exemplo de um ponto de ônibus substituído por um candidato válido.**



**Figura 5. Rota de ônibus mais segura para uma determinada linha na cidade de São Paulo considerando possíveis modificações.**

#### 4. Resultados e Análises

No desenvolvimento deste trabalho foram utilizadas duas máquinas. Para construir grande parte da proposta apresentada, foi utilizado um notebook Lenovo Ideapad 330S com sistema operacional Linux Mint 20 Cinamon, processador i5-8250U, 8GB de RAM e 240GB de SSD. Apesar deste computador ser capaz de executar todas as etapas necessárias, uma delas se mostrou computacionalmente intensa: a obtenção de todas as possibilidades de rotas entre pontos de parada da rota.

Como forma de diminuir o tempo de processamento necessário, foi utilizado o serviço de processamento “Cluster DCT” disponibilizado pela Diretoria de Tecnologia da Informação da Universidade Federal de Viçosa (DTI/UFV) [DTI/UFV]. Para este trabalho foi utilizada a alocação de um único núcleo e 3GB de RAM. Com isso, foi necessário gerar um *script* que executasse essa parte em específico e salvasse o progresso em um arquivo, para que os resultados pudessem ser utilizados localmente depois.

Uma vez que o trabalho possui algumas etapas custosas computacionalmente, foi utilizada uma amostra significativa do conjunto completo de rotas de ônibus. Para isso, foram escolhidas aleatoriamente 200 linhas dentre as 2.183 disponíveis. Estas foram selecionadas com base na distribuição uniforme discreta. Isto é válido pois garante que não sejam geradas amostras tendenciosas que beneficiem os resultados.

Com o objetivo de avaliar as principais características das soluções de roteamento rodoviário seguro desenvolvidas neste trabalho, bem como para realizar uma série de comparações, os seguintes cenários foram implementados:

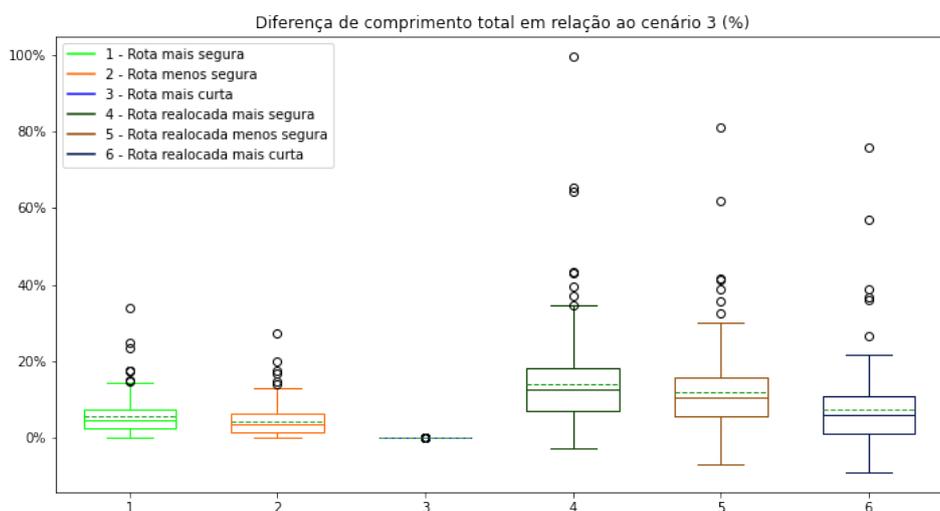
- *Cenário 1* - Rotas com pontos de paradas originais e trajeto entre eles com deslocamento mais seguro.
- *Cenário 2* - Rotas com pontos de paradas originais e trajeto entre eles com deslocamento menos seguro.
- *Cenário 3* - Rotas com pontos de paradas originais e trajeto entre eles com deslocamento mais curto. Este cenário, em particular, representa a forma como as linhas de ônibus operam atualmente.
- *Cenário 4* - Rotas com pontos de paradas realocados e trajeto entre eles com deslocamento mais seguro.
- *Cenário 5* - Rotas com pontos de paradas realocados e trajeto entre eles com deslocamento menos seguro.
- *Cenário 6* - Rotas com pontos de paradas realocados e trajeto entre eles com deslocamento mais curto.

Com os cenários formalmente definidos, foram estabelecidas as métricas necessárias para avaliação dos resultados ligados aos aspectos de segurança e distância percorrida gerados pelo trabalho. Todas as métricas foram coletadas para os seis cenários e os resultados comparados como forma de se estabelecer o compromisso entre os dois aspectos mencionados acima para cada um deles, bem como entre os mesmos. Vale ressaltar que os cenários 2 e 5 servem apenas como parâmetros de comparação com os outros, visto que representam extremos de insegurança. Logo, não são relevantes para uma possível implantação em serviços de transporte.

Antes de apresentar as métricas, vale ressaltar que os resultados obtidos e que serão apresentados posteriormente foram adquiridos com base em análises relativas ao cenário 3. Como brevemente apresentado na definição dos cenários, o cenário 3 foi escolhido como referência para a forma atual de como as linhas de ônibus operam dentro da cidade. Esta decisão foi tomada uma vez que o *dataset* de linhas de ônibus possui apenas os pontos de parada. Foi necessário traçar rotas entre eles para gerar soluções iniciais, sendo a do cenário 3 aquela mais coerente para ocupar esta posição, justamente por apresentar o caminho mais curto entre dois pontos de parada.

“Comprimento Total Percorrido” é a primeira métrica apresentada e avalia o comprimento total das rotas completas. Ela é calculada em relação ao cenário 3. Em outras palavras, mostra a quantidade total em metros percorridos a mais para cada cenário, entre o ponto de origem e o ponto de destino, em relação ao cenário 3. A pergunta a ser respondida pela análise dessa métrica é: criando soluções alternativas é possível manter um comprimento próximo ao original? Na Figura 6 percebe-se que houve um aumento médio de aproximadamente 10% se comparado o cenário 1 com o 3. Mais precisamente, o comprimento médio para o cenário 3 é 18.683,48m e para o cenário 1 é 19.736,19m. No

gráfico também se encontram os cenários 4 a 6, os quais possuem as paradas realocadas. Percebe-se que todos apresentam um aumento na distância total. Uma das causas disso é que em várias paradas de diversas linhas vai haver realocação e que essas, por sua vez, em geral aumentam a distância necessária a ser percorrida, ainda que haja limite para a distância na mudança. Sendo assim, percebe-se também que para o cenário 4 é possível notar que já há um aumento de aproximadamente 13%, tendo em média uma distância de 21.200,11m. Outro cenário relevante para este contexto é o cenário 6, observa-se que esse também recebe um acréscimo de aproximadamente 5% a 10%, que em média representa 19.918,59m.



**Figura 6. Diferença da distância percorrida em relação ao cenário 3.**

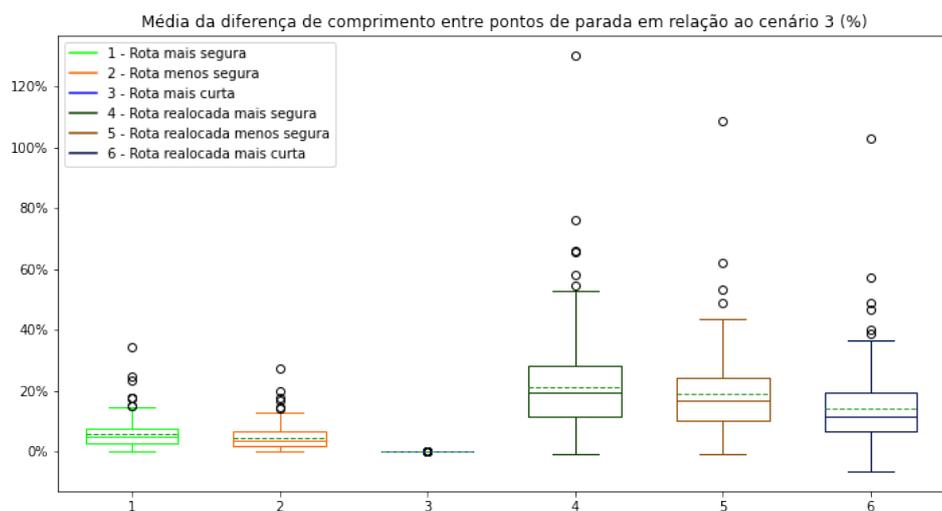
Em paralelo ao que foi apresentado acima, “Comprimento Médio entre Pontos de Parada” é uma métrica que representa a distância média que o motorista terá que percorrer entre dois pontos de parada adjacentes.

Diante disso, na Figura 7 o cenário 3 também é escolhido como referência, possuindo uma distância média entre pontos de parada de 504,63m. Para o cenário 1 nota-se que em média há um aumento de aproximadamente 10%, totalizando 531,40m. Já para o cenário 4 há um aumento mais significativo, por volta de 20%, contabilizando 601,12m. Assim como na métrica anterior, neste caso também é relevante observar o cenário 6. Observa-se que esse também recebe um acréscimo de aproximadamente 10%, que em média representa 566,69m.

As duas métricas apresentadas acima mostraram o impacto que os critérios de segurança causaram nas distâncias percorridas, não destacando a análise de segurança em si. Desta forma, a próxima métrica já é um exemplo do quesito segurança.

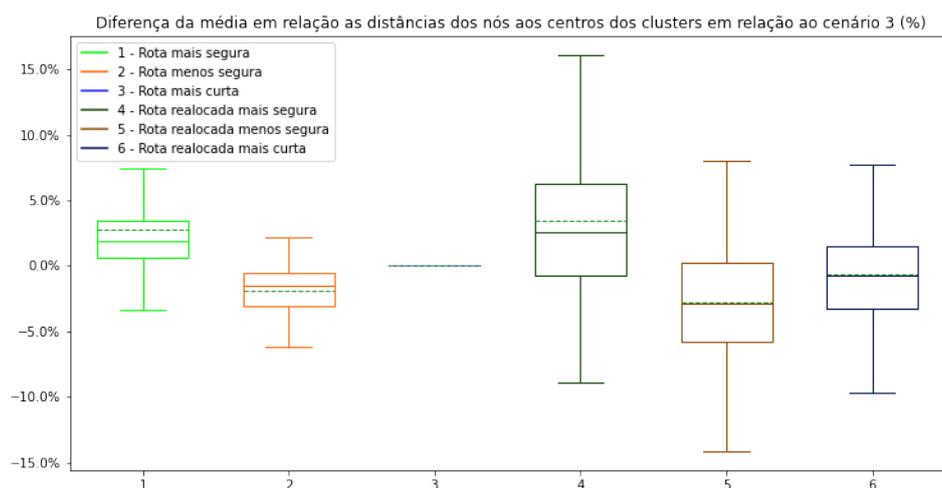
Como já foi mencionado anteriormente, as rotas são construídas com bases nos vértices do grafo, que consequentemente representam as esquinas da cidade analisada. Com isso, cada vértice serve como referência tanto para a rota quanto para a segurança, pois através dele é possível checar o quão próximo cada parte da rota está de uma região perigosa.

Sendo assim, “Média das Distâncias dos Nós em Relação aos seus Respectiveiros



**Figura 7. Diferença da distância percorrida entre pontos de parada em relação ao cenário 3.**

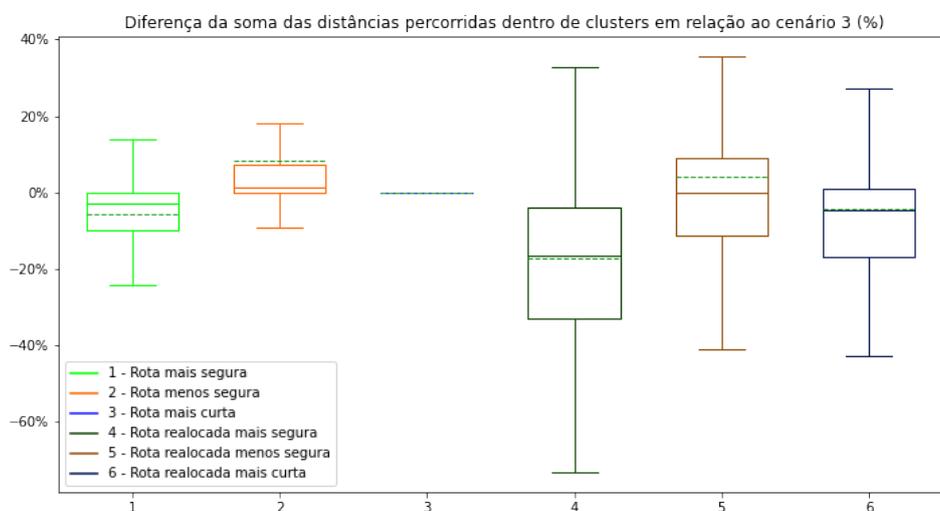
Centros de *Clusters* mais Próximos” é uma métrica utilizada para identificar, em média, o quão próximo ao centro de uma região perigosa o ônibus está passando. Na Figura 8 percebe-se que tanto o cenário 1 quanto o cenário 4 possuem em média uma melhoria de 3% se comparado com o cenário 3. Em média, a distância real nesta análise é 384,95m para o cenário 1, 384,84m para o cenário 4 e 373,93m para o cenário 3. Além disso, através do gráfico é possível notar que o cenário 4 possui um conjunto de valores médios mais distintos entre si. Isso pode ser notado observando-se a distância entre os quartis exibidos. Com isso, percebe-se que há vértices, neste cenário, com uma distância 5% maior se comparados aos originais.



**Figura 8. Diferença em relação ao cenário 3 das distâncias de cada esquina das rotas ao centro de cluster mais próximo.**

“Distância percorrida dentro de *clusters* criminais” é uma métrica relacionada ao quanto um ônibus trafegaria dentro de *clusters* criminais durante o trajeto. Para isso foram feitas algumas adaptações. Como não é possível identificar as intercessões entre vias e

*clusters* na ferramenta *OSMNx*, foi considerado que para um trecho estar dentro de uma região perigosa ele deve começar e terminar em uma esquina localizada dentro de um *cluster*. Com isso, diante da rota completa, cada par de esquinas que satisfizerem estas condições serão contabilizadas para esta métrica. A Figura 9 também compara os demais cenários com o cenário 3. De acordo com essa figura, percebe-se que os cenários mais seguros percorrem uma distância menor dentro de regiões perigosas, por isso a porcentagem para estes cenários é negativa. Para o cenário 1 tem-se um ganho médio de 5%, já para o cenário 4, há um ganho maior, por volta de 20%. Para o contexto analisado, essas diferenças podem ser representadas pelas distâncias calculadas, que são: 3.089,33m para o cenário 3, 2.919,66m para o cenário 1 e 2.508,39m para o cenário 4.



**Figura 9. Diferenças em relação ao cenário 3 das distâncias percorridas dentro de regiões perigosas.**

Por fim, “Análise da Realocação de Pontos de Parada” é uma métrica que faz um comparativo entre os pontos de paradas originais e os novos propostos. Além das diferenças apresentadas anteriormente, também é possível analisar e comparar a quantidade de pontos de paradas dentro de *clusters* criminais antes e depois das realocações sugeridas. Antes de apresentar essa comparação é preciso ressaltar que diante dos seis cenários apresentados, os resultados são os mesmos para os cenários 1 a 3, assim como também os mesmos valores atribuídos aos cenários 4 a 6. Sendo assim, na Tabela 3 é possível notar que com a implementação da possibilidade de realocação de pontos de parada, vários pontos originais foram retirados de regiões inseguras e realocados para regiões fora de *clusters* criminais. Além disso, à título de comparação, a média de pontos de paradas em regiões com *clusters* criminais é de aproximadamente 4,03. Se comparada à média de pontos de paradas por rota, cujo valor é 42,85, percebe-se que a solução que aborda o conceito de realocação de pontos de paradas, contém, em média, apenas 10% dos pontos em regiões perigosas.

## 5. Conclusão e Trabalhos Futuros

O trabalho apresentado neste artigo descreveu a aplicação e, posterior ampliação, de uma solução da literatura para elaboração de rotas seguras no contexto de rotas rodoviárias

<b>Categoria</b>	<b>Mínimo</b>	<b>Mediana</b>	<b>Máximo</b>	<b>Média</b>	<b>Desvio padrão</b>
Pontos de parada originais	0	10	36	11,15	7,1
Pontos de parada realocados	0	3,5	17	4,03	3,29

**Tabela 3. Estatísticas de pontos de parada localizados em *clusters* criminais para rotas completas.**

na cidade de São Paulo. Percebe-se que a principal questão levantada, referente à possibilidade de criar alternativas de rotas mais seguras para transportes coletivos, levando em consideração a questão social, foi alcançada. Isso pôde ser constatado nos resultados que apresentaram melhorias nos níveis de segurança das linhas e dos passageiros sem impactar significativamente nos desvios e distâncias percorridas. Em destaque, é possível observar que o cenário 1, que mantém os pontos de paradas oficiais, foi capaz de oferecer alternativas mais seguras com baixo acréscimo na distância percorrida. Por outro lado, o cenário 4 apresenta uma proposta que modifica os pontos de paradas originais, mas em compensação oferece rotas e pontos de paradas mais seguros, também com baixo acréscimo na distância percorrida.

Como trabalhos futuros existem alguns caminhos relevantes para a continuidade do trabalho apresentado. O primeiro deles é avaliar o impacto do uso de rotas seguras nos ônibus junto à população, levando em consideração seu real padrão de mobilidade. Outro fator interessante que pode ser adicionado à proposta apresentada está relacionado ao aspecto temporal, ou seja, avaliar se em diferentes horários do dia o comportamento de segurança ainda se mantém o mesmo. Por fim, pode também ser avaliado o impacto do uso de outros algoritmos, bem como seus valores de parâmetros, no agrupamento de dados criminais para identificação de regiões inseguras da cidade. Ou utilizar estratégias que envolvem ponderamento de arestas do grafo para determinar as regiões inseguras.

## Referências

- Atzori, L., Iera, A., and Morabito, G. (2016). Understanding the internet of things: definition, potentials, and societal role of a fast evolving paradigm. *Ad Hoc Networks*, 56.
- Atzori, L., Lera, A., and Morabito, G. (2018). The internet of things: A survey.
- 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.
- DataWorld. The cloud-native data catalog. and so much more... Disponível em: <https://data.world/>. Acesso em: 30 de abril de 2021.
- DBSCAN. How dbscan works and why should we use it? Disponível em: <https://towardsdatascience.com/how-dbscan-works-and-why-should-i-use-it-443b4a191c80>. Acesso em: 30 de abril de 2021.
- DTI/UFV. Dct - uso do ambiente. Disponível em: <https://dct.ufv.br/uso-do-ambiente/>. Acesso em: 4 de maio 2021.

- Félix Mata, Miguel Torres-Ruiz, G. G. R. Q. R. Z.-F. M. M.-I. E. L. (2016). A mobile information system based on crowd-sensed and official crime data for finding safe routes: A case study of Mexico City. *Mobile Information Systems*, 2016:1–11.
- Galbrun, E., Pelechris, K., and Terzi, E. (2016). Urban navigation beyond shortest route: The case of safe paths. *Information Systems*, 57:160–171.
- Here. Location for developers. Disponível em: <https://developer.here.com/>. Acesso em: 12 de maio 2021.
- InterSCity. Enabling the future internet for smart cities. Disponível em: <https://interscity.org/>. Acesso em: 30 de abril de 2021.
- Ladeira, L., de Souza, A., Silva, T., Filho, G. R., Peixoto, M. M., and Villas, L. (2020). Cerva: Roteamento contextual para veículos com risco espaço-temporal. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 379–392, Porto Alegre, RS, Brasil. SBC.
- Ladeira, L., Souza, A., Pereira, G., Silva, T. H., and Villas, L. (2019). Serviço de sugestão de rotas seguras para veículos. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 608–621, Porto Alegre, RS, Brasil. SBC.
- Levy, S., Xiong, W., Belding, E., and Wang, W. Y. (2020). Saferoute: Learning to navigate streets safely in an urban environment. *ACM Trans. Intell. Syst. Technol.*, 11(6).
- Liu, Q., Kumar, S., and Mago, V. (2017). Safernet: Safe transportation routing in the era of internet of vehicles and mobile crowd sensing. *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)*.
- OSMnx. Osmnx 1.0.1. Disponível em: <https://osmnx.readthedocs.io/en/stable/>. Acesso em: 30 de abril de 2021.
- Santos, F. A., Rodrigues, D. O., Silva, T. H., Loureiro, A. A. F., and Villas, L. A. (2017). Rotas veiculares cientes de contexto: Arcabouço e análise usando dados oficiais e sensorizados por usuários sobre crimes. In *Anais do XXII Workshop de Gerência e Operação de Redes e Serviços*, Porto Alegre, RS, Brasil. SBC.
- Tompson, L., Partridge, H., and Shepherd, N. (2009). Hot routes: Developing a new technique for the spatial analysis of crime. *Crime Mapping: A Journal of Research and Practice*, 1(1):77–96.
- Verbraeken, J., Wolting, M., Katzy, J., Kloppenburg, J., Verbelen, T., and Rellermeyer, J. S. (2020). A survey on distributed machine learning. *ACM Comput. Surv.*, 53(2).
- Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., and Chen, C. (2011). Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1624–1639.