

VINÍCIUS GABRIEL DE JESUS ALMEIDA

**ROTAS ALTERNATIVAS CENTRADAS EM USUÁRIOS
PARA VEÍCULOS COLETIVOS COM
SEGURANÇA E DISTÂNCIA EQUILIBRADOS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientadora: Thais R. M. Braga Silva

Coorientador: Fabrício A. Silva

**VIÇOSA - MINAS GERAIS
2024**

Ficha catalográfica elaborada pela Biblioteca da Universidade Federal de Viçosa - Campus Florestal

T

A447r
2024
Almeida, Vinícius Gabriel de Jesus, 1998-
Rotas alternativas centradas em usuários para veículos coletivos com segurança e distância equilibrados / Vinícius Gabriel de Jesus Almeida. – Florestal, MG, 2024.
1 dissertação eletrônica (83 f.): il. (algumas color.).

Orientador: Thais Regina de Moura Braga Silva.
Dissertação (mestrado) - Universidade Federal de Viçosa, Instituto de Ciências Exatas e Tecnológicas, 2024.
Referências bibliográficas: f. 79-83.
DOI: <https://doi.org/10.47328/ufvcaf.2024.013>
Modo de acesso: World Wide Web.

1. Transporte urbano. 2. Levantamento de rotas. 3. Sistemas inteligentes de veículos rodoviários. I. Silva, Thais Regina de Moura Braga, 1982-. II. Universidade Federal de Viçosa. Instituto de Ciências Exatas e Tecnológicas. Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDD 23. ed. 623.8933


VINÍCIUS GABRIEL DE JESUS ALMEIDA

**ROTAS ALTERNATIVAS CENTRADAS EM USUÁRIOS
PARA VEÍCULOS COLETIVOS COM
SEGURANÇA E DISTÂNCIA EQUILIBRADOS**


Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

APROVADA: 19 de março de 2024.

Assentimento:

Documento assinado digitalmente
 VINICIUS GABRIEL DE JESUS ALMEIDA
Data: 04/08/2024 11:34:54-0300
Verifique em <https://validar.iti.gov.br>

Vinícius Gabriel de Jesus Almeida
Autor

Documento assinado digitalmente
 THAIS REGINA DE MOURA BRAGA SILVA
Data: 05/08/2024 10:13:04-0300
Verifique em <https://validar.iti.gov.br>

Thais R. M. Braga Silva
Orientadora

AGRADECIMENTOS

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Esta entidade, em grande parte do trabalho, financiou o desenvolvimento da pesquisa com apoio financeiro.

Agradeço especialmente à minha orientadora, Thais R. M. B. Silva, que esteve comigo desde 2018, sempre me dando suporte e trabalhando em conjunto em diversos projetos, além de também sempre me apoiar e incentivar a seguir na área acadêmica. Meus agradecimentos também ao meu coorientador, Fabrício A. Silva, que sempre me proporcionou suporte e incentivo ao longo dessa trajetória.

Meus agradecimentos estendem-se a Marcos Tozatti da Silva (Funcionário da área de desenvolvimento, inovação e criação da URBS - Urbanização de Curitiba S/A) e Luciana Durand Garda (Chefe de Gabinete da SPTrans), que contribuíram com os dados públicos de mobilidade urbana que nos faltavam e que eram essenciais para o término do trabalho. Também expresso minha gratidão à empresa Cinnecta por nos fornecer os dados públicos do censo já processados, o que agilizou o processo de análise de resultados.

Agradeço também aos integrantes do grupo de estudos NESPeD e ao corpo docente da UFV (Florestal/Viçosa), que também contribuíram com a minha caminhada acadêmica e formação.

Agradeço aos meus pais, Helenita Maria de Jesus Almeida e Nivaldo Geraldo de Almeida, que sempre me apoiaram e incentivaram a estudar, nunca me deixando desistir e seguindo comigo nessa caminhada até o final. Também expresso minha gratidão à minha namorada Júlia Costa de Faria por estar ao meu lado, ajudando, incentivando e apoiando-me desde os primeiros anos da graduação. Por fim, agradeço aos meus amigos, que também sempre me ajudaram e apoiaram de diversas formas ao longo dessa jornada.

RESUMO

ALMEIDA, Vinícius Gabriel de Jesus, M.Sc., Universidade Federal de Viçosa, março de 2024. **Rotas Alternativas Centradas em Usuários para Veículos Coletivos com Segurança e Distância Equilibrados.** Orientadora: Thais R. M. Braga Silva. Coorientador: Fabrício A. Silva.

Com o aumento da disponibilidade de dados e dispositivos conectados nos centros urbanos, diferentes aplicações desenvolvidas para apoiar a construção de cidades inteligentes que fazem uso de informações contextuais estão ganhando espaço e se tornando mais comuns. Contextos são dados sobre pessoas, lugares ou objetos e os Sistemas de Transporte Inteligentes (STIs), em particular, fazem uso destes com o objetivo de aprimorar a mobilidade dos cidadãos. Neste recorte, um aspecto que vem ganhando destaque é a construção de rotas ciente de contexto. Este tipo de solução utiliza um ou mais tipos de contexto para elaborar a trajetória de agentes da cidade. Dentre os possíveis contextos a serem utilizados, a segurança é um que vem sendo bastante investigado recentemente. No entanto, embora diferentes trabalhos já tenham implementado e avaliado soluções de rotas seguras, a maioria foi testada para veículos particulares e avaliada sob o ponto de vista computacional. Este trabalho apresenta RACIONAL, um modelo genérico de cálculo de rotas seguras adaptado às peculiaridades dos coletivos urbanos. Isso inclui a necessidade de restringir as modificações propostas nas trajetórias, respeitando a cobertura da rota, e garantindo a acessibilidade aos passageiros que fazem uso e dependem deste meio de transporte. Resultados obtidos com estudos de caso realizados para dois municípios distintos, São Paulo e Curitiba, mostraram que o modelo é capaz de lidar com diferentes perfis de cidades, proporcionando soluções viáveis com um ganho significativo em segurança, associado a um baixo acréscimo de deslocamento, tanto pelos ônibus em suas rotas como dos passageiros no acesso a pontos de parada alternativos.

Palavras-chave: Mobilidade Urbana; Rotas Ciente de Contexto; Transporte Público e Segurança

ABSTRACT

ALMEIDA, Vinícius Gabriel de Jesus, M.Sc., Universidade Federal de Viçosa, March, 2024. **User-Centric Alternative Routes for Collective Vehicles with Balanced Safety and Distance**. Advisor: Thais R. M. Braga Silva. Co-advisor: Fabrício A. Silva.

With the increased availability of data and connected devices in urban centers, various applications developed to support the creation of smart cities using contextual information are gaining traction and becoming more common. Contexts are data about people, places, or objects, and Intelligent Transportation Systems (ITS) in particular make use of these to enhance citizens' mobility. In this context, a significant emerging aspect is the construction of context-aware routes. This type of solution utilizes one or more types of context to determine the trajectories of city agents. Among the possible contexts to be used, security has been a focus of recent investigations. However, although different studies have implemented and evaluated secure route solutions, most have been tested for private vehicles and assessed from a computational perspective. This work presents RACIONAL, a generic model for calculating secure routes adapted to the peculiarities of urban public transportation. This includes the need to limit proposed modifications in trajectories, maintaining route coverage, and ensuring accessibility for passengers who use and depend on this mode of transport. Case studies conducted in two different municipalities, São Paulo and Curitiba, demonstrated that the model can handle various city profiles, providing viable solutions with significant safety gains and a minimal increase in displacement for both buses on their routes and passengers accessing alternative bus stops.

Keywords: Urban Mobility; Context-Aware Routes; Public Transportation and Safety

LISTA DE FIGURAS

3.1	O modelo RACIONAL e seus componentes.	20
4.1	Exemplo de quadriculas no mapa do Brasil	39
4.2	Mapa de relevo de São Paulo	40
4.3	Mapa de relevo de Curitiba	40
4.4	Análise dos polígonos criminais gerados na cidade de São Paulo.	45
4.5	Análise dos polígonos criminais gerados na cidade de Curitiba.	46
4.6	Mapa de calor dos bairros que mais tiveram polígonos criminais na cidade de São Paulo.	47
4.7	Mapa de calor dos bairros que mais tiveram polígonos criminais na cidade de Curitiba.	48
5.1	Acréscimo de impedância nas sugestões de realocação.	56
5.2	Pontos de paradas removidos de dentro de <i>clusters</i> criminais.	57
5.3	Soma das distâncias sequenciais dentro dos mesmos <i>clusters</i> de São Paulo.	58
5.4	Soma das distâncias sequenciais dentro dos mesmos <i>clusters</i> de Curitiba.	59
5.5	Acréscimo total das rotas na cidade de São Paulo.	59
5.6	Acréscimo total das rotas na cidade de Curitiba.	60
5.7	Média dos acréscimos dos trechos de São Paulo.	61
5.8	Média dos acréscimos dos trechos de Curitiba.	62
5.9	Média das distâncias dos nós em relação aos centros dos <i>clusters</i> de São Paulo.	62
5.10	Média das distância dos nós em relação aos centros dos <i>clusters</i> de Curitiba.	63
5.11	Média das distâncias dos pontos de parada em relação aos centros dos <i>clusters</i> de São Paulo.	64
5.12	Média das distância dos pontos de parada em relação aos centros dos <i>clusters</i> de Curitiba.	64
5.13	Soma das distâncias sequenciais dentro dos mesmos <i>clusters</i> criminais para trechos com maiores fluxos de pessoas de São Paulo.	66
5.14	Soma das distâncias sequenciais dentro dos mesmos <i>clusters</i> criminais para trechos com maiores fluxos de pessoas de Curitiba.	66
5.15	Acréscimo total do trecho na cidade de São Paulo.	67
5.16	Acréscimo total do trecho na cidade de Curitiba.	67
5.17	Distância dos nós em relação aos centros dos <i>clusters</i> para a cidade de São Paulo.	68
5.18	Distância dos nós em relação aos centros dos <i>clusters</i> para a cidade de Curitiba.	69
5.19	Gráfico de pareto do agrupamento dos <i>clusters</i> criminais nos setores censitários para a cidade de São Paulo.	70
5.20	Gráfico de pareto do agrupamento dos <i>clusters</i> criminais nos setores censitários para a cidade de Curitiba.	71
5.21	Gráfico de pareto do agrupamento dos pontos de paradas avaliados que foram realocados na cidade de São Paulo.	72

5.22	Gráfico de pareto do agrupamento dos pontos de paradas avaliados que foram realocados na cidade de Curitiba.	73
5.23	Mapa de calor dos bairros que mais receberam realocações na cidade de São Paulo.	75
5.24	Mapa de calor dos bairros que mais receberam realocações na cidade de Curitiba.	76

LISTA DE TABELAS

2.1	Características do trabalho proposto em relação ao estado da arte. . . .	18
4.1	<i>Dataset</i> criminal limpo da cidade de São Paulo	35
4.2	<i>Dataset</i> criminal limpo da cidade de Curitiba	36
4.3	<i>Dataset</i> bruto das linhas de ônibus de São Paulo.	37
4.4	<i>Dataset</i> bruto dos pontos de parada de São Paulo.	37
4.5	<i>Dataset</i> bruto das linhas de ônibus de Curitiba.	38
4.6	<i>Dataset</i> bruto dos pontos de parada de Curitiba.	38
4.7	<i>Dataset</i> com dados brutos das altitudes (m).	39
4.8	<i>Dataset</i> com dados de fluxos de passageiros entre pares de pontos de parada.	41
4.9	<i>Dataset</i> simplificado com coluna de total de fluxo do dia.	41
4.10	Informações complementares do campo <i>stop_id</i>	42
4.11	<i>Dataset</i> de mobilidade da cidade de Curitiba.	42
4.12	<i>Dataset</i> de mobilidade da cidade de Curitiba enriquecido com inferência.	43
4.13	<i>Dataset</i> final com todas informações.	49
4.14	<i>Dataset</i> final com todas informações.	50
4.15	Linhas de ônibus prontas para o modelo.	50
5.1	Tabela com as métricas de impacto dos setores, causadas pelas realocações dos pontos de parada na cidade de São Paulo.	73
5.2	Tabela com as métricas de impacto dos setores, causadas pelas realocações dos pontos de parada na cidade de Curitiba.	74

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Problema e Hipótese	11
1.2	Justificativa	12
1.3	Objetivo Geral e Objetivos Específicos	12
1.4	Organização do Trabalho	13
2	TRABALHOS RELACIONADOS	14
3	RACIONAL: ROTAS ALTERNATIVAS CENTRADAS EM USUÁRIOS PARA VEÍCULOS COLETIVOS COM SEGURANÇA E DISTÂNCIA EQUILIBRADOS	19
3.1	Visão Geral do Modelo RACIONAL	19
3.2	Preparação da Base de Dados	20
3.2.1	Dados do mapa viário	21
3.2.2	Dados criminais	21
3.2.3	Dados de ônibus	22
3.2.4	Dados de relevo	22
3.2.5	Dados de mobilidade	22
3.3	Construção de Abstrações	23
3.3.1	Clusterização dos dados criminais	23
3.3.2	Construção do grafo enriquecido da cidade	23
3.4	Construção de Rotas Seguras	24
3.4.1	Definição das Rotas	24
3.4.2	Construção de todas as rotas possíveis com pontos de parada originais	25
3.4.3	Construção de todas as rotas possíveis utilizando pontos de parada alternativos	26
3.5	Avaliação	29
3.5.1	Filtragem das rotas construídas	29
3.5.2	Cálculos de métricas por linha	30
3.5.3	Cálculos de métricas por trecho	30
4	ESTUDOS DE CASO	32
4.1	Materiais	32
4.2	Preparação da Base de Dados	34
4.2.1	Escolha das cidades e obtenção das bases de dados	34
4.2.2	Dados do mapa viário	34
4.2.3	Dados criminais	35
4.2.4	Dados de ônibus	36
4.2.5	Dados de relevo	39
4.2.6	Dados de mobilidade	40
4.3	Construção de Abstrações	43
4.3.1	Clusterização dos dados criminais	43
4.3.2	Construção do grafo enriquecido da cidade	47
4.4	Construção das Rotas Seguras	50

4.4.1	Construção de todas as rotas possíveis com pontos de parada originais	50
4.4.2	Construção de todas as rotas possíveis utilizando pontos de parada alternativos	51
4.5	Avaliação	53
4.5.1	Filtragem das rotas construídas	53
4.5.2	Cálculos de métricas por linha	54
4.5.3	Cálculo de métricas por trecho	54
5	RESULTADOS	55
5.1	Avaliando as rotas completas	55
5.2	Cálculo de métricas por trecho	63
5.3	Avaliação da solução proposta com base em fatores socioeconômicos	68
5.3.1	Avaliando a distribuição socioeconômica dos polígonos criminais das cidades	69
5.3.2	Avaliando os pontos de paradas e realocações feitas	71
6	CONCLUSÃO	77
	REFERÊNCIAS BIBLIOGRÁFICAS	79

Capítulo 1

Introdução

A *Mobilidade Urbana* é um conceito relacionado a facilidade de deslocamento de indivíduos, como pedestres, ciclistas, usuários de transportes coletivos, motoristas e outros na área urbana [Aragão \(2006\)](#). Ela desempenha um papel crucial em diversas atividades sociais e econômicas, incluindo estudo, trabalho e lazer [Magagnin and da Silva \(2008\)](#). No entanto, diante da crescente complexidade dos centros urbanos, inúmeros desafios ligados ao deslocamento de seus agentes puderam ser observados. A partir da década de 1990, com a implementação de uma nova política de atração de investimentos para a indústria automobilística, a capacidade de produção de automóveis e motocicletas mais que triplicou. Essa expansão foi impulsionada por incentivos políticos, como a redução de cargas tributárias, e o aumento da renda das famílias, levando a um significativo aumento no número de veículos particulares, conforme destacado em [Carvalho \(2016\)](#). Esses e outros fatores aumentaram a complexidade da mobilidade urbana, impactando negativamente a saúde física e mental dos habitantes de grandes cidades. Uma das soluções possíveis para se mitigar este problema é através do uso de *Sistemas de Transporte Inteligente (STI)*.

Os STIs têm como objetivo, entre outras coisas, aprimorar a mobilidade, o fluxo dos transportes e a segurança [Qureshi and Abdullah \(2013\)](#). Além disso, visam aumentar a produtividade das pessoas e contribuir para a redução dos efeitos nocivos do trânsito [Cunha et al. \(2017\)](#). Para alcançar essas melhorias, é essencial a integração de tecnologias de computação e comunicação nos veículos e na infraestrutura da cidade. É importante destacar também que os STIs não se limitam apenas às melhorias nas condições de tráfego de veículos, pois também atuam no propósito de tornar o setor de transportes mais sustentável, eficiente e seguro, conforme destacado por [Cunha et al. \(2017\)](#). Um dos pilares para o STI é a chamada *Mobilidade Ciente de Contexto*, definida como o uso de tipos de dados distintos sobre lugares, pessoas e objetos, chamados contextos, para oferecer melhorias aos fluxos de veículos, motoristas e passageiros de transportes coletivos [Vahdat-Nejad et al. \(2016\)](#).

Um das áreas envolvidas com a promoção da Mobilidade Ciente de Contexto é a *Construção de Rotas Cientes de Contexto*. Ela está relacionada ao uso de dados de um ou mais tipos de contextos na determinação do caminho a ser seguido. Assim, os

usuários podem escolher o tipo de trajeto desejado, seja ele o mais curto [Madkour et al. \(2017\)](#), o menos congestionado [Pires et al. \(2014\)](#), o mais cênico [Denstadli and Jacobsen \(2011\)](#), o mais seguro [Ladeira et al. \(2019\)](#), entre outros [Santos et al. \(2017\)](#). Dentre os contextos apresentados, a segurança tem despertado bastante interesse na literatura sobre Mobilidade Urbana nos últimos anos.

Rotas seguras são aquelas que evitam áreas com alto grau de criminalidade. Ou seja, a partir de dados contextuais criminais, que posteriormente seriam convertidos para uma estratégia de identificação de regiões inseguras, o algoritmo de recomendação de rotas seria capaz de entender esse contexto e, assim, construir o trajeto de modo a se distanciar dessas regiões. Esse tipo de contexto de rotas pode ser utilizado por diferentes agentes de uma cidade, como pedestres ([Boarnet et al. \(2005\)](#) e [Bura et al. \(2019\)](#)), ciclistas ([Winters et al. \(2012\)](#)), veículos particulares ([Ladeira et al. \(2019\)](#)), e também por veículos públicos, como ônibus urbanos, por exemplo. Para o caso específico de veículos públicos, como os ônibus urbanos, surge a seguinte particularidade: esses veículos operam em linhas com pontos de parada subsequentes, e qualquer modificação no trajeto deve considerar as regiões que possuem paradas obrigatórias, as quais podem estar localizadas em áreas inseguras. Portanto, as alterações podem envolver a mudança nos pontos de parada.

1.1 Problema e Hipótese

A construção de rotas seguras para ônibus pode trazer benefícios aos passageiros do sistema de transporte municipal, como: diminuir a exposição destes a áreas inseguras, de modo geral; evitar que estes esperem seus ônibus em pontos de paradas localizados em regiões com altos índices criminais; e também proteger a integridade física dos passageiros em situações de conflito que possam surgir durante o trajeto, comprometendo a segurança do ônibus e, conseqüentemente, dos próprios passageiros. Porém, a modificação dos trajetos, e possivelmente dos pontos de parada, poderá ter um impacto significativo em relação à nova distância total percorrida pelos veículos, bem como no conforto ao acesso à este sistema por seus usuários. Desta forma, por se tratar de uma rota que prioriza a segurança, os caminhos resultantes podem adicionar trajetos com comprimentos maiores em comparação ao cenário que prioriza a distância percorrida pelos veículos. Em relação à perspectiva do passageiro, há a possibilidade de enfrentarem problemas de acesso às novas sugestões de pontos de parada, que podem estar em locais distantes, inacessíveis e exigir muito esforço extra para serem alcançados. Desse modo, uma pergunta de pesquisa que surge é: seria possível se obter uma solução de construção de rotas seguras para veículos coletivos capaz de equilibrar segurança, distância e conforto dos usuários?

A hipótese considerada neste trabalho é a de que é possível a proposição de uma solução, que possa ser aplicada inclusive para diferentes cidades, capaz de melhorar a segurança dos passageiros por meio da alteração de trajetos e pontos de parada, desde que obedecendo a restrições e limites relacionados ao esforço necessário para que o passageiro alcance os locais de embarque e desembarque.

Por fim, é importante mencionar que, devido à natureza do problema e da solução proposta para resolvê-lo, algumas questões éticas podem surgir. A partir do momento em que é necessário estabelecer critérios para determinar se uma área é considerada segura, como ajustes nos parâmetros para a distribuição de *clusters*, essas manipulações podem, por si só, introduzir viés ao trabalho.

1.2 Justificativa

O desenvolvimento deste trabalho é coerente e significativo, uma vez que o transporte coletivo desempenha um papel crucial para milhões de pessoas em todo o Brasil. Para o ano de 2023, a média de viagens diárias foi de aproximadamente 32,1 milhões de passageiros ¹. Em contrapartida, é um setor frequentemente impactado por problemas de segurança, pois segundo um levantamento da NTU em 2016 ², foram registrados 53.479 assaltos em ônibus nos últimos 19 anos. Esses dados evidenciam que os passageiros estão frequentemente expostos a um alto risco de enfrentar situações criminosas durante suas viagens.

Além disso, não existem trabalhos na literatura, até onde pudemos investigar, que proponham soluções para construção de rotas seguras para ônibus, em particular considerando a questão do conforto ao acesso por parte dos passageiros.

1.3 Objetivo Geral e Objetivos Específicos

O objetivo geral deste trabalho é elaborar e avaliar uma solução centrada em usuário e genérica, i.e., passível de ser utilizada em diferentes cidades, que possibilite ao gestor municipal do transporte coletivo público identificar e utilizar rotas mais seguras para os ônibus, com ou sem mudança dos pontos de parada existentes.

Os objetivos específicos a serem desenvolvidos como forma de alcançar o objetivo geral são:

- Organizar a solução proposta na forma de um modelo de fácil compreensão,

¹Associação Nacional das Empresas de Transportes Urbanos (NTU): Os grandes números da mobilidade urbana. <https://www.ntu.org.br/novo/AreasInternas.aspx?idArea=7>

²Associação Nacional das Empresas de Transportes Urbanos (NTU): Um levantamento realizado pela NTU em 2016 registrou um total de 53.479 assaltos em ônibus nos últimos 19 anos. <https://www.ntu.org.br/novo/NoticiaCompleta.aspx?idArea=10&idNoticia=1106>

customizável e aplicável de forma consistente para diferentes cidades, produzindo como saída as novas rotas a serem seguidas pelas linhas de ônibus, seus benefícios em termos de segurança e o custo em relação à distância e conforto dos usuários.

- Propor e avaliar algoritmos da literatura a serem utilizados para a clusterização de crimes e construção das rotas seguras.
- Propor uma solução original, centrada em usuário, para realocação dos pontos de parada para locais mais seguros das linhas do sistema de transporte coletivo. Vale ressaltar que, além dos aspectos técnicos, essa solução também visa preservar a qualidade de vida e o conforto da população usuária do transporte coletivo.
- Elaborar novas métricas para avaliação dos ganhos em termos de segurança das rotas seguras apontadas pelo modelo e quanto ao impacto social das mudanças sugeridas.
- Avaliar o modelo proposto aplicando-o à duas cidades com características em termos de tamanho, magnitude do sistema de transporte, relevo e criminalidade diversas.

1.4 Organização do Trabalho

Sendo assim, o restante desta dissertação está organizado da seguinte forma. O Capítulo 2 conta com uma visão geral do estado da arte atual, além dos trabalhos relacionados que contribuíram com o projeto. No Capítulo 3, é apresentado o modelo RACIONAL desenvolvido neste trabalho, contando com detalhes de como foi projetado o modelo genérico responsável por construir rotas seguras para veículos coletivos urbanos. Já no Capítulo 4, é realizado o teste da implementação do modelo desenvolvido. Neste estudo de caso, são escolhidas duas cidades, que possuem dados suficientes para a aplicação do modelo, e nelas são aplicados todos os passos necessários para se obter as rotas mais seguras. Para validar o resultado da solução o Capítulo 5 conta com os principais resultados obtidos, junto com as observações e peculiaridades encontradas para cada cidade. Desta forma, também é possível avaliar diferentes situações e como o modelo se comportou para cada uma delas. Por fim, o Capítulo 6 apresenta as conclusões e trabalhos futuros, além de uma síntese dos resultados, reforçando a viabilidade que a utilização do modelo pode trazer para contribuir com a melhoria dos transportes coletivos.

Capítulo 2

Trabalhos relacionados

Sistemas de Transporte Inteligentes (STIs) são parte de um campo inovador que interage em diversas áreas do transporte em cidades, como gerenciamento, controle, infraestrutura, operações, políticas e métodos de controle, entre outros. Por isso, são capazes de desempenhar um papel vital no mundo [Qureshi and Abdullah \(2013\)](#). Entre seus benefícios, destacam-se a capacidade de reduzir o congestionamento de tráfego, diminuir o número de acidentes e oferecer vantagens como melhoria no fluxo de mobilidade, controle de poluição, prevenção de acidentes e aumento da confiabilidade no tempo de viagem [Garg and Kaur \(2022\)](#). Por se tratar de uma área já conhecida e amplamente desenvolvida, existem diversas subáreas dentro dos STIs [Qureshi and Abdullah \(2013\)](#), que incluem:

- Sistemas de gerenciamento de estradas arteriais e rodovias, com trabalhos como [Gao and Liu \(2013\)](#), [Dia and Thomas \(2011\)](#), [Bachmann et al. \(2013\)](#), entre outros.
- Sistemas de gerenciamento de frete, apresentando estudos como [Poon et al. \(2009\)](#), [Crainic et al. \(2009\)](#), [Zumerchik et al. \(2009\)](#), entre outros.
- Sistemas de gerenciamento de trânsito (TMS), com pesquisas como [Yan et al. \(2012\)](#), [Diab and El-Geneidy \(2012\)](#), [HONGYANGA \(2013\)](#), entre outros.
- Sistemas de gerenciamento de incidentes, incluindo estudos como [Kim et al. \(2007\)](#), [Wang et al. \(2012\)](#), [Shah et al. \(2008\)](#), entre outros.
- Sistemas de gerenciamento de emergência, com publicações como [Lambert et al. \(2013\)](#), [Ozguven and Ozbay \(2013\)](#), [Yoon et al. \(2008\)](#), entre outros.
- Sistemas Regionais de Informação Multimodal e de Viajantes/Gestão de Informação (IM), que englobam trabalhos como [Natvig and Vennesland \(2010\)](#), [Zhang et al. \(2011\)](#), [Peng \(1997\)](#), entre outros.

Deste modo, a fim de possibilitar a implementação dos STIs, existe a Mobilidade Ciente de Contexto, que está situada como um pilar importante da área, assim como brevemente apresentado no Capítulo 1. Se tratando de contextos, conceito

fundamental em ambiente de computação pervasivo, estes podem ser definidos como: qualquer informação que possa ser utilizada para caracterizar a situação de uma entidade. Uma entidade é uma pessoa, lugar ou objeto que é considerado relevante para a interação entre um usuário e uma aplicação, incluindo o próprio usuário e as aplicações (Dey (2001), Abowd et al. (1999)). Um sistema é sensível ao contexto se usa contexto para fornecer informações e/ou serviços relevantes ao usuário, onde a relevância depende da tarefa em si. Em geral, o desenvolvimento de qualquer aplicação sensível ao contexto consiste nos seguintes três componentes principais (Vahdat-Nejad (2014), Loke (2006)) que são:

1. Aquisição de contexto: coleta de informações contextuais por sensores.
2. Processamento: emprego de técnicas de raciocínio para obter informações contextuais de alto nível (como as atividades do usuário).
3. Atuação: prestar serviços ao usuário de acordo com sua situação atual.

No estudo de Vahdat-Nejad et al. (2016), é possível identificar exemplos de áreas de aplicação para soluções fundamentadas em mobilidade ciente de contexto. Nesse contexto, são destacados trabalhos que utilizam redes ad-hoc veiculares (VANETs). Essas redes permitem que um conjunto de veículos troque informações dinamicamente e sem fio enquanto se deslocam Cavalcante et al. (2012). Entre as possíveis aplicações, encontram-se estudos que abordam temas como:

- Alerta de colisão Al-Sultan et al. (2014).
- Notificação de condições perigosas na estrada Naja (2013).
- Aviso sobre a aproximação de veículos de emergência Al-Sultan et al. (2014).
- Assistência para ultrapassagem Olaverri-Monreal et al. (2010).
- Notificação sobre as condições de tráfego Wedel et al. (2009) e Bauza and Gozávez (2013).
- Identificação de pontos de interesse, como estacionamentos, postos de gasolina, hotéis ou restaurantes Naja (2013), dentre outras possibilidades.

Ao focar na área de Construção de Rotas Conscientes de Contexto dentro da Mobilidade Consciente de Contexto, diversos estudos na literatura têm apresentado soluções que buscam fornecer rotas calculadas com base em critérios que vão além da métrica de distância, considerando fatores como índices de acidentes e registros de criminalidade. Algumas dessas soluções foram inclusive destacadas no Capítulo 1. Portanto, neste trabalho, que tem como objetivo implementar uma

solução baseada em rotas conscientes de contexto, diferente da abordagem convencional focada apenas no menor tempo e distância, foram selecionados estudos que exploram esse tipo de solução alternativa na literatura. Além disso, devido à natureza da solução que envolve o uso de transporte público, foram realizadas buscas por trabalhos relacionados aos veículos coletivos urbanos para compreender o estado atual das soluções disponíveis. Assim, ao final deste capítulo, será possível entender como as soluções levantadas se relacionam e se alinham com o propósito deste trabalho, que consiste em desenvolver rotas seguras e centradas no usuário para veículos coletivos urbanos.

A seguir, serão apresentados os principais trabalhos encontrados e as relações e contribuições que estes apresentaram para o desenvolvimento da solução proposta.

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, e colisões e condições das vias para propor variações de rotas ao decorrer do dia. CERVA [Ladeira et al. \(2020\)](#) é uma solução para oferecer alternativas de rotas baseada em clusterização KDE para um dado contexto 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 referentes à cidade de Chicago, incluindo informações criminais oficiais e condições de tráfego provenientes de APIs de sistemas reais, como o *Here Maps*. VTq [de Souza and Villas \(2020\)](#) é um sistema completo para planejamento de rotas com foco em prever a dinâmica futura de fatores urbanos, tais como mobilidade e criminalidade. Os autores utilizaram na solução uma rede neural e um algoritmo de aprendizado por reforço.

Existem muitos estudos que apresentam propostas para confecção de rotas para veículos particulares, considerando especialmente o índice de segurança das mesmas. [Félix Mata et al. \(2016\)](#) apresentam um *framework*, construído através de redes bayesianas, que elabora rotas seguras com base em um sistema de recomendação híbrido que utiliza dados oficiais e do *Twitter*. O trabalho obteve cerca de 75% de precisão nas recomendações utilizando apenas os dados da rede social. Além disso, também apresenta flexibilidade quanto a 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 Filadélfia. 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.

Em [Ladeira et al. \(2019\)](#), o algoritmo DBSCAN é utilizado para a clusterização de crimes obtidos para a cidade de Chicago. Considerando as rotas mais frequentes

seguidas pelos motoristas desta localidade, os autores descrevem uma solução, chamada função *Parzen Window*, capaz de construir rotas seguras escolhendo, entre a origem e o destino, as vias mais afastadas dos *clusters* criminais encontrados. Tais rotas são ainda modificadas em janelas temporais que consideram intervalos de horários ao longo do dia. A proposta de [Ladeira et al. \(2019\)](#) é utilizada como base para a construção do modelo de construção de rotas seguras proposto neste trabalho. Apesar de se tratar de uma solução com foco em veículos particulares e utilizar janelas temporais com intervalos curtos, a estratégia apresentada pelos autores opta por um formato de clusterização que preserva a distribuição original dos crimes, sendo ainda capaz de identificar e ponderar o nível de segurança de cada parte de uma rota por meio de uma única equação compacta, porém efetiva. Além disso, o trabalho escolhido sintetiza de forma satisfatória as escolhas mais recorrentes e melhor avaliadas pelos demais artigos relacionados encontrados na literatura.

O estudo realizado por [Oliveira et al. \(2019\)](#) investiga os fatores que contribuem para o alto índice de assaltos a ônibus na cidade de Belo Horizonte, Brasil. As principais áreas de interesse da pesquisa abordam os padrões de assaltos, a distribuição espacial desses incidentes, os locais mais suscetíveis a roubos e as características ambientais associadas a eles. Embora o estudo não proponha soluções específicas para rotas seguras, sua conclusão enfatiza que a promoção da segurança durante as viagens deve se concentrar na redução das oportunidades de crimes. Portanto, destaca-se a importância de um esforço colaborativo entre agências governamentais e indivíduos para atingir esse objetivo. Por fim, [Tompson et al. \(2009\)](#) apresentam *Hot Routes*, uma proposta também voltada para ônibus. A solução, baseada em clusterização KDE, tem como principal objetivo identificar regiões inseguras, referentes à criminalidade, como pontos de paradas ou trechos do percurso, na cidade de Londres e exibi-las em um mapa. Esta proposta foi a única encontrada que foca em linhas de ônibus. Entretanto, diferente do presente trabalho, não propõe novas sugestões de rotas para substituir as identificadas como inseguras. Além disso, não possui nenhum método para alterar os pontos de parada para tornar a rota mais segura.

A Tabela 2.1 apresenta uma comparação do estado da arte com a proposta neste trabalho.

	Características			
	Projetado para	Algoritmo	Janela Temporal	Contexto
RACIONAL	Ônibus	DBSCAN e Parzen Window	Não	Criminal
CERVA 2020	Veículos Particulares	KDE	Sim	Criminal, Acidentes e Mobilidade
VTq 2020	Veículos Particulares	LSTM	Sim	Criminal e Mobilidade
Solução base 2019	Veículos Particulares	DBSCAN e Parzen Window	Sim	Criminal
SafeRNet 2017	Veículos Particulares	Redes Bayesianas	Sim	Trânsito (Acidentes)
A Mobile Information System 2016	Veículos Particulares	Redes Bayesianas	Sim	Criminal
SafePaths 2016	Veículos Particulares	KDE	Não	Criminal
Hot Routes 2009	Ônibus	KDE	Não	Criminal

Tabela 2.1: Características do trabalho proposto em relação ao estado da arte.

Capítulo 3

RACIONAL: Rotas Alternativas Centradas em usuários para veículos coletivos com segurança e distância equilibrados

Neste capítulo, será inicialmente apresentado o modelo RACIONAL, proposto para possibilitar a identificação de rotas seguras, sejam elas com seus pontos de paradas originais ou então com a opção de realocação de pontos de parada para regiões externas de *clusters* criminais. Essa segunda opção conta com uma análise centrada no usuário que visa evitar uma má distribuição destes pontos de parada, de modo que se torne inacessível ou exija muito esforço extra, seja físico ou de deslocamento, para o novo ponto de parada. Inicialmente será feita uma descrição geral deste modelo, destacando-se seus componentes e a relevância de cada um deles. Em seguida, cada componente será detalhado, de forma a mostrar quais são os itens esperados ou tarefas a serem realizadas e seus resultados, de maneira que seja possível observar como o modelo pode ser utilizado para diferentes cidades.

3.1 Visão Geral do Modelo RACIONAL

Como pode ser observado na Figura 3.1, o modelo RACIONAL é dividido em quatro componentes, cada um deles essencial para o funcionamento da solução como um todo. No primeiro, **Preparação da Base de Dados**, o principal objetivo é garantir uma base de dados sólida, contendo todas as informações necessárias para o fluxo subsequente. Isso é feito apresentando e justificando a obrigatoriedade de certas informações e explicando por que elas precisam estar disponíveis posteriormente.

No componente seguinte, **Construção de Abstrações**, os dados brutos são processados e padronizados para tornarem-se disponíveis para o modelo. Nessa etapa, destaca-se a construção da clusterização dos dados criminais e a representação da cidade-alvo por meio de uma abstração, centralizando as informações extraídas

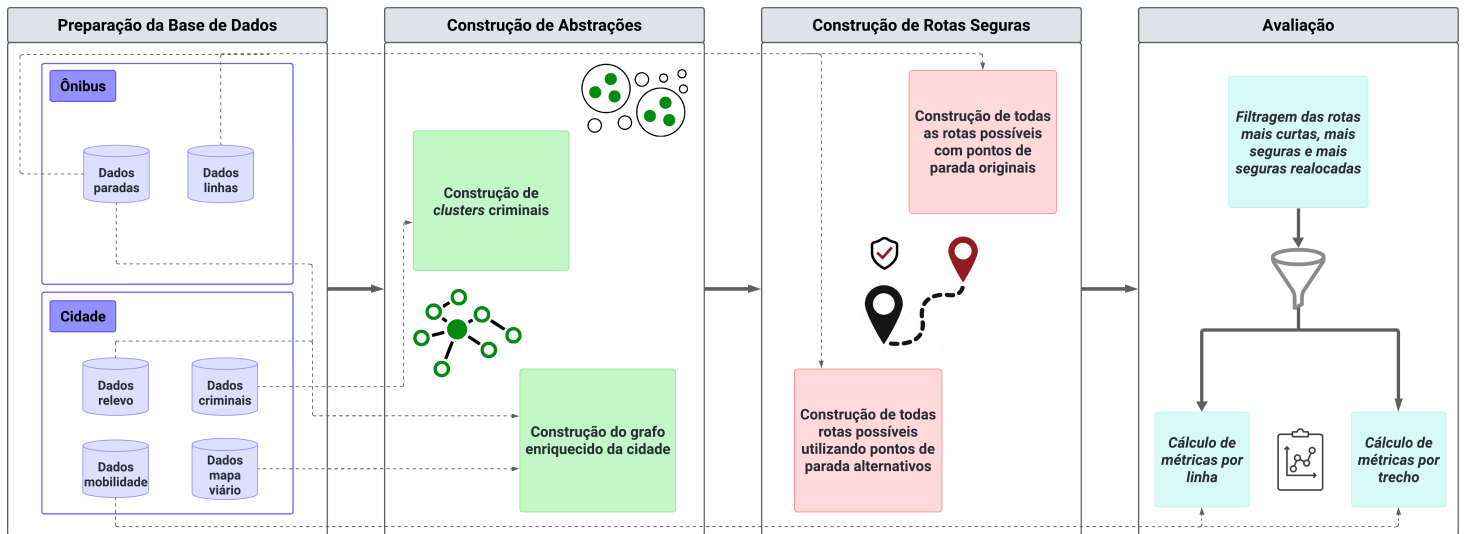


Figura 3.1: O modelo RACIONAL e seus componentes.

das bases de dados.

No terceiro componente, **Construção de Rotas Seguras**, a solução é de fato executada. Portanto, esta seção detalha como gerar as rotas seguras, os critérios, funções e algoritmos aplicados no processo, além de todas as manipulações extras necessárias para permitir o uso de pontos de parada alternativos.

Por fim, temos a **Avaliação**, que detalha o processo de filtragem das rotas construídas, visando extrair apenas as mais seguras, tanto para pontos de paradas originais quanto para aqueles que envolvem pontos de parada alternativos. Em seguida, inicia-se o cálculo das métricas, marcando o encerramento da aplicação do modelo. Portanto, percebe-se que esta etapa está mais voltada para avaliar o desempenho e a viabilidade da solução desenvolvida.

3.2 Preparação da Base de Dados

Para que o modelo seja executado de forma eficiente, é importante garantir que a etapa de preparação da base de dados seja feita de maneira coerente, assegurando que a obtenção dos dados básicos e obrigatórios ocorra de modo que todas as etapas subsequentes tenham informações precisas para sua execução.

Deste modo, uma divisão estrutural é representada na Figura 3.1. Observa-se que existem dois grandes grupos: ônibus e cidade. Essa separação existe para mostrar em qual área da solução os *datasets* vão impactar. Por exemplo, para ônibus, os dados de linhas e dados de paradas são essenciais para construir a representação da forma como esses veículos operam dentro da cidade. Já os dados criminais, apesar de serem dados do contexto urbano, no modelo RACIONAL, o impacto desses está atrelado à forma como as rotas seguras são construídas, sendo fatores determinantes

para cada parte do trajeto selecionado. Sendo assim, para o grupo da cidade, informações como dados de relevo e dados do mapa viário são essenciais para permitir tanto a representação física da cidade quanto para construir a análise centrada no usuário proposta, sendo determinantes para validar se as alternativas encontradas nas rotas estão alinhadas com o que se espera para os passageiros. Por fim, os dados de mobilidade são úteis no processo de entender como o fluxo de pessoas ocorre na cidade, além de contribuir para a análise dos resultados obtidos por meio de métricas.

Desta forma, é possível notar que o principal impeditivo para a replicação deste modelo é a ausência dos *datasets* básicos necessários para construir a solução. Portanto, é importante garantir que o local escolhido como estudo de caso contenha pelo menos esses dados disponíveis. Sendo assim, a seguir, serão apresentados maiores detalhes para a obtenção desses.

3.2.1 Dados do mapa viário

O primeiro passo para iniciar o uso do modelo RACIONAL é obter os dados do mapa viário da cidade escolhida. Como algumas informações adicionais sobre as cidades estão contidas nas outras categorias de dados apresentadas, neste momento, são essenciais apenas os dados que representam as vias (ruas) e as conexões feitas entre elas, representadas por esquinas. Esses dados são de suma importância para a construção da abstração da cidade, presente no componente seguinte.

3.2.2 Dados criminais

Os dados criminais também são obrigatórios no modelo, pois sem eles não é possível aplicar a etapa posterior de clusterização dos crimes. Para o modelo em si, é necessário que estes crimes tenham algum tipo de descrição do ocorrido, para que seja possível filtrá-los de maneira a permanecer apenas com aqueles que sejam relevantes para o contexto de rotas avaliado. Neste trabalho, por exemplo, sugere-se selecionar crimes relacionados com furtos e roubos, mais especificamente voltados para pedestres e veículos. Quanto maior a relação dos crimes filtrados com ameaças que de fato perturbam a segurança de usuários do sistema de transportes coletivos, melhor a solução final de rotas seguras apontadas pelo modelo. Por fim, é necessário ter algum tipo de informação do local da ocorrência do crime, de preferência sendo um dado geolocalizado, ou com uma aproximação boa que possa ser convertida para geolocalização posteriormente. Este dado é essencial, pois sem ele não é possível construir os *clusters* criminais mencionados, e conseqüentemente, as rotas almejadas.

3.2.3 Dados de ônibus

Como a aplicação do modelo é sob o contexto de veículos coletivos urbanos, que neste caso, se tratam de ônibus municipais, é necessário conhecer os trajetos feitos por esses. Idealmente, esta base de dados deverá estar completa, com todas as linhas de ônibus existentes na cidade, já possuindo a informação da rota que o veículo faz e com seus respectivos pontos de paradas. Entretanto, como é comum a dificuldade em se obter um conjunto abrangente de dados, neste modelo requer-se no mínimo que se tenha as seguintes informações: a geolocalização dos pontos de paradas, necessárias tanto para construir as linhas de ônibus, quanto para analisar todo o fluxo de mobilidade dos passageiros; e a segunda, as linhas de ônibus, que são representadas como sequências de pontos de paradas.

3.2.4 Dados de relevo

Os dados de relevo são importantes, pois possibilitam avaliar o ganho ou perda de altitude de um local para outro. Essas informações são essenciais no modelo RACIONAL para verificar se as sugestões de realocação de pontos de parada não exigem esforço físico excessivo dos passageiros, evitando que eles enfrentem aclives acentuados, por exemplo. Espera-se que as referências de altitude não sejam muito espaçadas, pois, caso contrário, todos os lugares avaliados terão uma variação nula de elevação, o que pode prejudicar os resultados e rotas geradas. Com isso, uma estrutura de dados que contenha geolocalização, seguida da altitude (em metros) do local, é suficiente para realizar a análise proposta.

3.2.5 Dados de mobilidade

O objetivo do uso dos dados de mobilidade neste trabalho está relacionado à identificação dos fluxos mais percorridos pelos cidadãos nas cidades. No entanto, por se tratar de uma solução que avalia alternativas de rotas para veículos coletivos urbanos, o fluxo de mobilidade ideal é aquele que representa o embarque e desembarque de passageiros pela cidade. Esses dados são essenciais para permitirem a identificação dos principais trechos das rotas, ou seja, aqueles que recebem o maior fluxo de pessoas.

Contudo, devido à dificuldade de obter essa informação de forma precisa, é também viável realizar essa análise de maneira aproximada. Isso pode ser alcançado por meio de dados de mobilidade de pedestres, uma vez que é suficiente ter o registro de partida e destino para realizar o recorte do trecho com o maior fluxo de pessoas.

Assim, pode-se observar que este dado é utilizado para compreender e validar os resultados gerados, não sendo essencial para a construção das rotas propriamente ditas. Dessa forma, torna-se opcional, especialmente se o objetivo for apenas construir as rotas com o modelo, sem a necessidade de uma análise mais aprofundada.

3.3 Construção de Abstrações

3.3.1 Clusterização dos dados criminais

Os dados criminais individuais obtidos para a base de dados não são efetivos para a solução proposta pelo modelo RACIONAL. Como são sempre registros únicos, e estão espalhados por toda a cidade, eles não funcionam para ponderar qual região está mais concentrada em crimes ou não, o que dificulta a identificação precisa das regiões inseguras de fato.

Para isso, algoritmos de clusterização são relevantes para criar esses agrupamentos e evidenciar tais regiões. Para este tipo de problema tratado nessa solução, existe apenas uma ressalva. Como se trata de manipulações em cima das rotas percorridas pelos ônibus, e também da manipulação de pontos de paradas que envolvem passageiros, os *clusters* devem de alguma forma ficar próximos da localização dos seus crimes originais que foram agrupados, pois as modificações propostas pelo modelo são feitas em raios de ações restritos e que envolvem outros fatores também. Desta forma, a única restrição é utilizar algum algoritmo de clusterização que respeite essa condição. Um dos algoritmos que respeita esta restrição, o qual foi utilizado nos estudos de caso deste trabalho, foi o *Density Based Spatial Clustering of Application with Noise* (DBSCAN) ¹. Este algoritmo é interessante pois constrói o agrupamento com base na proximidade dos pontos vizinhos que estão sendo avaliados, além de permitir escolher também um limiar mínimo de ocorrências necessárias para que o *cluster* resultante seja válido. Com isso, ao final desse processo, o agrupamento gerado fica na posição original inicial, onde os pontos selecionados estão localizados.

3.3.2 Construção do grafo enriquecido da cidade

Nesta etapa, embora já possuamos os dados apresentados na Subseção 3.2.1, que representam a cidade, é crucial construir uma abstração que a represente de maneira operável. Isso implica no uso de uma estrutura de dados robusta que contribua para

¹DBSCAN: Método clássico de agrupamento baseado em densidade. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>

o modelo, permitindo que essas informações sejam carregadas de forma a ficarem disponíveis para simular o deslocamento dentro da cidade posteriormente.

Assim, após a construção dessa abstração, é necessário enriquecê-la com dados complementares, agrupando todas as informações úteis em uma única estrutura que as gerencie. Dessa forma, é necessário mapear tanto os dados que representam os pontos de parada quanto as informações de relevo. O cenário ideal é realizar esse mapeamento nas localizações originais, mantendo o restante da estrutura intacta. Contudo, caso isso não seja possível, uma alternativa é realizar o mapeamento de forma aproximada, considerando as esquinas mais próximas.

No contexto deste trabalho, é necessário o uso de estruturas de dados que contenham algum tipo de grafo, uma vez que seu uso facilita o relacionamento entre as ruas e esquinas, sendo facilmente mapeadas como arestas e vértices, respectivamente. Além disso, tais estruturas oferecem soluções já conhecidas para a construção de caminhos entre dois vértices quaisquer, como os algoritmos de *Dijkstra* e *Bellman–Ford*, por exemplo.

3.4 Construção de Rotas Seguras

3.4.1 Definição das Rotas

Com o intuito de avaliar posteriormente as rotas seguras alternativas propostas pelo modelo RACIONAL, 3 cenários distintos serão considerados. No primeiro cenário, chamado de **ROTAS MAIS CURTAS**, são construídas as trajetórias mais curtas a partir dos pontos de paradas das linhas já utilizadas pelos ônibus, conforme descritas pelos dados obtidos na (seção 3.2.3), uma vez que não se conhece o trajeto oficial. Esta escolha permite simular um trajeto justo e próximo da realidade, além de possibilitar a criação de um cenário de comparação para avaliar posteriormente os possíveis acréscimos gerados pela solução. Caso os dados sobre as rotas dos ônibus não estejam completos, isto é, não apresentem a trajetória entre os pontos de parada, uma solução alternativa é criar as mesmas utilizando os pontos de paradas geolocalizados de forma sequencial e então construindo o caminho mais curto entre os pares sequenciais.

A partir da definição da rota mais curta, utilizada como cenário atual da cidade, torna-se possível estabelecer um parâmetro de comparação para avaliar a solução RACIONAL desenvolvida e observar em quais aspectos o trabalho contribuirá para melhorias nas cidades. Com isso, o primeiro cenário de rotas que contemplam a abordagem de segurança das cidades são as **ROTAS MAIS SEGURAS**. Essas são obtidas através de métricas de segurança, que serão detalhadas a seguir, tendo como garantia que a rota selecionada para a linha avaliada sempre será a com maior

segurança possível, independente do deslocamento necessário para percorrê-la. Perceba que para que esse cenário possua bons resultados, é essencial que diversas rotas alternativas que interligam os pares de pontos de paradas sequenciais das linhas sejam geradas e avaliadas, com o intuito de criar uma boa base de rotas possíveis, para então escolher aquela com o melhor índice de segurança.

Por fim, o cenário de **ROTAS MAIS SEGURAS REALOCADAS**, que parte do mesmo princípio das rotas mais seguras, porém, com a possibilidade de permitir que pontos de paradas localizados dentro de polígonos criminais sejam realocados para outros pontos de paradas, desde que esses não estejam localizados dentro de polígonos criminais, e que respeitem os requisitos de realocação, que também serão detalhados a seguir.

3.4.2 Construção de todas as rotas possíveis com pontos de parada originais

A solução para construção de rotas seguras do modelo RACIONAL foi desenvolvida sobre os fundamentos teóricos encontrados no trabalho [Ladeira et al. \(2019\)](#), cujo enfoque inicial foi direcionado a veículos particulares. Entretanto, elementos da solução base, como a própria construção das rotas, inicialmente destinadas a veículos particulares neste trabalho, precisaram ser adaptados. Isso ocorre porque, ao impactar somente o motorista e os demais passageiros que possivelmente estão indo para o mesmo destino, essas rotas apresentam maior flexibilidade para propor mudanças de trajeto. Assim, ao adaptar a solução para o contexto do transporte público, com diversos interesses envolvidos e locais obrigatórios para se passar, as variações de rotas precisaram ser restritas para não descaracterizar o trajeto original. Contudo, apesar dessas limitações encontradas, o uso da solução foi adaptado, permitindo replicar o algoritmo de construção de rotas seguras, juntamente com a equação base de segurança, e política de clusterização utilizada. A última precisou apenas de ajustes nos parâmetros, uma vez que variam de acordo com as cidades e bases criminais utilizadas.

Para auxiliar na descrição da solução, foi criado um conjunto de definições formais para representar cada conceito, possibilitando sua citação ao longo do trabalho.

Definições: Uma cidade possui N linhas de ônibus distintas. Cada linha de ônibus é definida como $L_i, 1 \leq i \leq N$. Uma linha (L_i) pode percorrer uma ou mais rotas entre sua origem e seu destino. R_i é o conjunto de todas as possíveis rotas R_i^x existentes para a linha L_i , que posteriormente são filtradas em rotas de interesse, que em particular, são: R_i^c a rota mais curta (i.e., aquela que percorre o caminho com a menor distância), R_i^s a rota mais segura (i.e., aquela que percorre o caminho com o menor índice de criminalidade), e R_i^r a rota mais segura realocada (i.e., aquela que

percorre o caminho com o menor índice de criminalidade e que permite a realocação de pontos de paradas). Cada rota R_i^x é composta por uma sequência de m vértices v_j , os quais neste trabalho representam esquinas por onde o veículo deve passar. Assim, $R_i^x = \{v_0, v_1, \dots, v_{(m-1)}\}$, em que m é o número de vértices que compõem a rota. Seja P_i o conjunto de pontos de parada p_z da linha L_i , onde p_z é um vértice no qual o coletivo permite embarque e desembarque de passageiros. Logo, $P_i \subset R_i^x \forall R_i^x \in R_i$. Em outras palavras, todos os pontos de parada de uma linha devem estar presentes em todas as possíveis rotas dessa linha. Por fim, um trecho, $R_{t(p_i, p_j)}$, é um subconjunto de vértices de uma rota que ligam o par de pontos de parada p_i e p_j .

Para a identificação do grau de periculosidade das rotas, a equação da solução base, [Ladeira et al. \(2019\)](#), foi utilizada (Equação 3.1). Esta leva em consideração o resultado da clusterização dos crimes, neste caso, representada como polígonos criminais. Todos os trechos da rota $R_{t(p_i, p_j)}$ recebem uma pontuação $k_{(p_i, p_j)}$, que quando somadas, representam o valor K , responsável por indicar o nível de periculosidade da rota. Desta forma, quanto menor for este valor, mais segura será a rota. Essa equação, denominada *Parzen Window*, é uma forma não paramétrica para estimar a função densidade de probabilidade de uma variável aleatória [Babu and Viswanath \(2008\)](#).

$$k_{(p_i, p_j)} = \frac{1}{m} \sum_{a=1}^m \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x_a - x}{2\sigma^2}\right) \quad (3.1)$$

Na Equação 3.1, m indica a quantidade de vértices (representando esquinas) do trecho da rota entre p_i e p_j ; σ é o desvio padrão da quantidade de crimes por *cluster*; x_a representa a distância de cada vértice em relação ao centro do polígono criminal mais próximo; e x representa a distância da borda do polígono criminal mais próximo do vértice avaliado em relação ao seu centro.

No cenário **ROTAS MAIS SEGURAS**, a rota mais segura é construída com base apenas no $k_{(p_i, p_j)}$ obtido para cada trecho. Desta forma, é escolhida, a cada trecho, a opção que contém o caminho com menor $k_{(p_i, p_j)}$.

O Algoritmo 1 formaliza como a rota R_i^s é calculada.

3.4.3 Construção de todas as rotas possíveis utilizando pontos de parada alternativos

No sistema de transporte coletivo, os pontos de paradas podem acabar sendo os locais onde os passageiros estão mais vulneráveis e suscetíveis a serem vítimas de crimes. No decorrer da construção da solução RACIONAL, apresentada neste trabalho, é possível observar que, independente das alternativas de rotas criadas, em algumas

Algoritmo 1: Identificação de R_i^s para a linha L_i

Data: P_i
Result: R_i^s

```

1  $R_i^s \leftarrow []$ ;
2 for  $index \leftarrow 0$  to  $(len(P_i) - 2)$  do
3    $origem \leftarrow P_i[index]$ ;
4    $destino \leftarrow P_i[index + 1]$ ;
5    $R_{t(origem,destino)} \leftarrow \mathbf{gera\_rotas}(origem,destino)$ ;
6    $niveis \leftarrow \langle K, R^x \rangle \leftarrow \{\}$ ; Tupla de níveis de crimes e rotas
7   for  $R_{t(origem,destino)}^x \in R_{t(origem,destino)}$  do
8      $niveis \leftarrow niveis + \{\langle \mathbf{parzen\_window}(R_{t(origem,destino)}^x), R_{t(origem,destino)}^x \rangle\}$ ;
9    $\mathbf{sort}(niveis)$ ;
10   $R_i^s \leftarrow R_i^s + niveis[0]$ ;
11 return  $R_i^s$ ;

```

situações, os próprios pontos de paradas seriam os locais menos seguros, pois mesmo que a rota contornasse a região através de um caminho mais seguro, a parada dentro de um *cluster* criminal seria inevitável.

Nesta seção é apresentada uma contribuição original do modelo RACIONAL que visa modificar os pontos de paradas das linhas, quando possível. No cenário **ROTAS MAIS SEGURAS**, as rotas criadas são geradas sempre respeitando os pontos de parada originais, alterando apenas os trajetos dos trechos de pontos de paradas consecutivos.

Quando um ponto de parada de uma linha está dentro de um *cluster* criminal, a segurança da rota é prejudicada, e acaba se tornando incoerente, pois desvia o trajeto mas a parada está dentro da região insegura. Desta forma, para contornar este tipo de situação, a estratégia utilizada para mitigar este problema é considerar a realocação do ponto de parada para outro próximo já existente, caso seja possível. A escolha desta alternativa foi feita para não interferir na infraestrutura da cidade, causando custos extras, além de também não gerar transtornos nas rotinas dos passageiros, que já conhecem os respectivos pontos de paradas existentes na sua região. Entretanto, a solução também pode ser modificada para aceitar a criação de novos pontos de paradas, bastando que a cidade que esteja implementando a solução esteja disposta a custear toda a infraestrutura extra necessária.

Para realizar as realocações de paradas, foram adotados principalmente critérios centrados em usuários, neste caso os usuários do transporte coletivo. Primeiramente, como brevemente mencionado, apenas pontos de paradas fisicamente disponíveis, que pertencem à uma linha qualquer, foram utilizados. Isso foi feito com o intuito de evitar a necessidade de intervenção das prefeituras das cidades para a criação de novos pontos. Com isso, para que um ponto de parada localizado dentro de

um polígono criminal seja removido, deve haver pelo menos algum outro ponto de parada fora de uma região insegura. Feito isso, a distância deste para o ponto original, em linha reta (distância euclidiana), deve possuir no máximo, *Raio* metros. Além disso, por se tratar de transporte terrestre, também foi considerada a distância real de locomoção pelas vias da cidade, medida em metros *Distancia*, que resultaria no possível deslocamento real que os passageiros deveriam percorrer a mais, no pior caso, para chegarem até o novo ponto de parada realocado. Dessa forma, esse valor deve possuir uma margem maior, visto que o deslocamento real dentro de cidades dificilmente será feito em linha reta.

Contudo, apesar das medidas citadas acima terem um impacto significativo para garantir que o ponto realocado fique nas redondezas do original, tentando diminuir o esforço dos passageiros para se deslocarem até o local novo, ele não é capaz de nos garantir que os passageiros não terão que lidar com altos aclives para chegar até o novo local. Por isso, para contornar essa situação e adotar um critério que de fato torne isso possível, foi utilizado o conceito de impedância. O trabalho de [Oliveira and Garcia \(2021\)](#), que é uma abordagem voltada para melhorar as condições de trabalhos de carrinheiros ao longo do Brasil, foi adaptado e utilizado como referência. A Equação 3.2, proposta por [Oliveira and Garcia \(2021\)](#) foi utilizada na solução RACIONAL para lidar com impedância, e conseqüentemente, diminuir o número de rotas com aclives.

$$i = \begin{cases} \theta^2 \cdot d, & \text{se } \theta > 0 \\ \theta \cdot (-1) \cdot d, & \text{caso contrário} \end{cases} \quad (3.2)$$

Na Equação 3.2, θ representa o ângulo da via e d , o tamanho da aresta (distância). A partir disso, tem-se um valor constante de peso para a aresta, que não precisa ser atualizado durante o percurso. A partir deste conceito, as simulações do trabalho são feitas aplicando esse peso da aresta nos esforços necessários, trabalho (W), que o carrinheiro teria que fazer no seu carrinho de coleta de material reciclável para concluir o deslocamento na aresta avaliada.

Entretanto, para o conceito de passageiros, essa abordagem se torna um pouco mais complexa, pois não há um corpo externo com um peso pré-definido, para que possamos calcular o trabalho final e usá-lo como referência. Por isso, neste caso, apenas essa mecânica básica não é suficiente, uma vez que a soma das forças internas de um corpo é sempre zero, e conseqüentemente, o trabalho é nulo, segundo o princípio da conservação de energia mecânica.

Com isso, a alternativa adotada foi obter este valor de impedância empiricamente e usá-lo como limiar de aceite ou recusa do ponto de parada candidato, e desconsiderar a métrica de trabalho (W) envolvida. Desta forma, a ideia por trás do experimento foi

obter as variáveis necessárias para se obter o valor da impedância i , que neste caso, se tratam de distância e inclinação da via. Sendo assim, a informação de distância é trivial de se conseguir, pois basta medir o quanto foi percorrido na via. Contudo, a inclinação requer um cuidado necessário para se obter. Para este trabalho, a ideia utilizada para se obter este ângulo foi o uso da aproximação do triângulo retângulo formado pela distância euclidiana percorrida (hipotenusa), e ganho de elevação ao percorrer esta via (cateto oposto). Com isso, através de trigonometria básica, o ângulo de inclinação pôde ser obtido. Logo, já com as duas variáveis necessárias, é possível obter a impedância i e então combinar e avaliar diferentes valores, até chegar em algum limiar coerente com a realidade da cidade alvo.

Por fim, com a aplicação de todas essas estratégias citadas anteriormente, é possível garantir que os passageiros não terão que percorrer distâncias maiores que os limites estabelecidos, e nem que terão que passar por acessos com aclives acentuados, além de evitar também o pior caso, ficar sem acesso ao novo ponto de parada. Desta forma, se algum dos critérios não for satisfeito, ou se o ponto de parada não estiver dentro de um *cluster* criminal, a realocação não é realizada. O Algoritmo 2 descreve todos esses passos.

Algoritmo 2: Realocação de Pontos de Parada de uma Linha L_i

Data: P_i

```

1  $num\_clusters \leftarrow \alpha, raio \leftarrow \beta, distancia \leftarrow \omega, impedancia \leftarrow \delta;$ 
2 for  $p_z$  in  $P_i$  do
3   if  $esta\_dentro\_cluster(p_z)$  then
4      $clusters\_proximos \leftarrow encontra\_clusters(p_z, num\_clusters);$ 
5      $p_{candidato} \leftarrow$ 
6       vertice\_fora\_cluster( $clusters\_proximos, raio, distancia, impedancia$ );
7     if  $p_{candidato}$  then
8        $p_z \leftarrow p_{candidato}$ 

```

3.5 Avaliação

3.5.1 Filtragem das rotas construídas

Como apresentado anteriormente, no cenário **ROTAS MAIS SEGURAS**, as rotas são obtidas através da comparação de todos os índices de periculosidade K , obtidos ao se aplicar a Equação 3.1. Com isso, ao testar todas as possibilidades de rotas possíveis para cada linha, é feito um ranqueamento, com o intuito de encontrar aquela que possui o menor índice K , sendo essa, a rota mais segura.

Para o cenário **ROTAS MAIS SEGURAS REALOCADAS** o processo também é feito da mesma forma, porém, um passo antes de construir essas rotas, é feito todo o procedimento apresentado para a realocação de pontos de paradas. Com isso, para o modelo, a estrutura continua a mesma, tendo apenas que construir alternativas de rotas para essas “novas linhas”, que já contêm os pontos realocados. Sendo assim, ao final do processo, diante das alternativas apresentadas para cada linha, o próximo passo é apenas ranqueá-las, assim como foi mencionado para o cenário **ROTAS MAIS SEGURAS**, e então se obter as rotas mais seguras realocadas.

3.5.2 Cálculos de métricas por linha

Nesta etapa, o modelo já terminou de construir todas as opções de rotas para cada cenário avaliado, diante das diferentes linhas submetidas ao processo. Com isso, vale ressaltar que esse passo é de apenas avaliação e validação dos resultados, para analisar se as alternativas de rotas sugeridas se comprometem com o propósito de segurança, e se sim, à qual custo de implementação na cidade alvo.

Dentre as opções de métricas para comparar e avaliar o desempenho das rotas sugeridas e o cenário base, existem diversas alternativas. Neste trabalho, alguns fatores foram escolhidos para realizar esta análise. Começando pelo principal fator, o de segurança, existem algumas considerações que devem ser feitas. Apesar do modelo sempre se basear no índice K para determinar a segurança da rota, esta métrica sozinha não é capaz de representar os reais ganhos obtidos. Isso se deve ao fato do valor obtido ser um número que possui grande parte de suas variações, quando se compara cenários, nas casas decimais mais baixas. Com isso, o uso desta informação não é muito recomendado, devido à dificuldade de visualização e interpretação. Sendo assim, outras formas de validar a segurança foram criadas, como a quantidade de pontos de paradas removidos de regiões inseguras e o distanciamento médio das rotas mais seguras em relação ao centro dos *clusters* mais próximos, dentre outras opções, que serão detalhadas e explicadas no capítulo de resultados (Capítulo 5).

3.5.3 Cálculos de métricas por trecho

Por fim, também é proposta uma métrica mais elaborada, que exige a base de dados de fluxo de mobilidade, para que possa ser construída. Por se tratar de um trabalho que constrói uma solução centrada no usuário, um dos cuidados necessários é, em diferentes aspectos, avaliar se esses não vão ser prejudicados. Para isso, é elaborada uma forma de checar os impactos das sugestões propostas. Apesar de já restringir os critérios de realocação dos pontos de paradas para viabilizar a migração dos passageiros, é importante validar se todas as modificações propostas são de fato

úteis, no contexto real. Ou seja, se os ganhos obtidos são em locais relevantes das rotas, onde mais pessoas poderiam ser mais impactadas.

Essa abordagem é tratada como trecho $R_{t(p_i, p_j)}$. O seu intuito, junto com o fator mobilidade, é identificar quais os principais trechos de cada rota quando o assunto é fluxo de passageiros. Com isso, são feitos recortes, extraíndo dessas rotas apenas esses intervalos principais, que possuem os maiores fluxos de mobilidade. Desta forma, a partir destes novos dados, as métricas já podem ser aplicadas normalmente, assim como foi feito no contexto anterior para as rotas completas R_j .

Capítulo 4

Estudos de Caso

4.1 Materiais

Para o desenvolvimento do RACIONAL, foram necessárias algumas tecnologias e ferramentas para auxiliar na construção, gerenciamento e obtenção dos resultados.

Como base do projeto, foi utilizada a linguagem de programação Python ¹, sendo a escolha dessa influenciada pela vasta gama de bibliotecas consolidadas para projetos que envolvam ciência de dados e *machine learning*, ponto essencial para construir uma base sólida para desenvolver a aplicação deste trabalho.

Para a organização do projeto, foi escolhido o sistema de controle de versões distribuído Git ², essencial para o gerenciamento e armazenamento dos avanços obtidos no projeto. A plataforma escolhida para uso em nuvem da ferramenta foi o GitLab ³.

Por se tratar de uma aplicação que demanda um grande volume de dados, além de estruturas de dados e códigos mais complexos, que exigem um uso significativo de memória e processamento, tornou-se necessário buscar ambientes de desenvolvimento que minimizassem ao máximo a perda de progresso durante a codificação. Com isso, para superar esse desafio, destaca-se o ambiente de desenvolvimento conhecido como Jupyter Notebook ⁴. Ele é capaz de salvar as etapas de um mesmo código utilizando o conceito de células, beneficiando-se do modo interativo do interpretador da linguagem. Dessa forma, é possível avançar no código, fazendo modificações apenas onde necessário, sem a necessidade de reexecutar as etapas que já foram validadas, o que reduz consideravelmente o tempo de desenvolvimento.

Outra ferramenta útil no processo de desenvolvimento e tratamento das bases de dados utilizadas é biblioteca Pandas ⁵. Ela oferece uma série de estruturas de dados

¹**Python:** <https://www.python.org/>

²**Git:** Is a free and open source distributed version control system.

³**GitLab:** Balance speed and security in a single platform <https://about.gitlab.com/>

⁴**Jupyter Notebook:** Web-based interactive computing platform. <https://jupyter.org/>

⁵**Pandas:** Biblioteca em Python para a manipulação e análise de dados. <https://pandas.pydata.org/>

e operações úteis na visualização, manipulação e limpeza de dados.

Para construir a clusterização utilizada no modelo, que será detalhada posteriormente, foi utilizada a biblioteca Scikit-learn ⁶. Essa é uma ferramenta consolidada quando se trata de aprendizado de máquina através de algoritmos supervisionados e não supervisionados em Python.

Todas as ferramentas citadas até o momento são muito populares para o desenvolvimento de projetos de ciência de dados em Python. Porém, por se tratar de um projeto que envolve manipulação de linhas de ônibus e clusterização de dados criminais, um estrutura mais complexa foi necessária. Com isso, de acordo com o contexto do projeto, o uso de uma estrutura de dados de grafo se torna a maneira mais compatível de representar e criar uma solução. Para isso, foi utilizada a biblioteca OSMnx ⁷, que se baseia na consolidada biblioteca NetworkX ⁸ para a construção dos grafos, além de enriquecer os vértices e arestas do grafo com dados do OpenStreetMap ⁹, que é um projeto de mapeamento colaborativo para criar um mapa livre e editável do mundo. Desta forma, com essa ferramenta, a estrutura do projeto já pode ser criada, convertendo as esquinas da cidade para vértices do grafo, e as ruas como arestas. A única limitação desta ferramenta está em relação à adição de vértices novos no grafo, não tornando possível adicionar pontos de paradas com suas localizações exatas no estrutura criada. Com isso, foi necessário fazer uma aproximação na representação dos pontos de parada da cidade. Isso fez com que cada ponto de parada fosse representado através da esquina localizada mais próxima dele. Portanto, apesar da principal limitação já ter sido apresentada e solucionada nesta etapa, os detalhes desta implementação serão apresentados a seguir nas próximas seções.

Por fim, para visualizar os resultados e estruturas de grafos criadas, foi utilizada a biblioteca Matplotlib ¹⁰, ferramenta também bastante difundida em projetos que utilizam a linguagem Python para construir análises e resultados visuais.

Para implementar e fazer uso de todas as ferramentas e tecnologias listadas acima, foi utilizado um notebook Lenovo Ideapad 330S, com sistema operacional Linux Mint 20 Cinnamon, processador i5-8250U, 20GB de RAM, 8GB de Swap, e 240GB de SSD.

⁶**Scikit-learn:** Biblioteca de aprendizado de máquina. <https://scikit-learn.org/stable/>

⁷**OSMnx:** Biblioteca de criação e manipulação de grafos. <https://osmnx.readthedocs.io/en/stable/>

⁸**NetworkX:** Biblioteca em Python para estudar grafos e redes. <https://networkx.org/>

⁹**OpenStreetMap:** Projeto de produção colaborativa de dados geo-espaciais abertos. <https://www.openstreetmap.org/>

¹⁰**Matplotlib:** Biblioteca para criação de gráficos e visualizações de dados em geral <https://matplotlib.org/>

4.2 Preparação da Base de Dados

O levantamento dos dados públicos junto com a disponibilidade dos mesmos foram pontos determinantes para a escolha das cidades utilizadas e avaliadas no trabalho. Nas seções a seguir, serão apresentadas com mais detalhes os motivos dessas escolhas, como estes dados foram obtidos, e as limpezas que foram necessárias para torná-los compatíveis com os objetivos do RACIONAL.

4.2.1 Escolha das cidades e obtenção das bases de dados

Para construir uma base de dados confiável e extensa, muitas cidades foram avaliadas e investigadas ao longo do processo de seleção. Apesar de muitas possuírem diferentes tipos de dados públicos de qualidade, ao fazermos o levantamento de todos os dados necessários para atender todas as etapas do projeto, muitas foram sendo descartadas, por não conterem em seus acervos dados essenciais que precisaríamos no futuro. Devido à essas limitações, a primeira cidade escolhida, São Paulo, foi uma das poucas que continham todos os dados necessários, a princípio.

Uma vez que para avaliar o aspecto genérico do modelo era fundamental que o mesmo pudesse ser aplicado em cidades diferentes, foi necessário selecionar um segundo município. Com isso, partimos do mesmo princípio de seleção da cidade de São Paulo, tentando achar bases de dados parecidas e coerentes com o modelo de projeto base. Desta forma, a segunda cidade escolhida foi Curitiba, devido à sua disponibilidade e qualidade de dados abertos.

4.2.2 Dados do mapa viário

As informações dos mapas viários das cidades podem ser obtidas de forma mais simples, pois estão contidas em uma das ferramentas utilizadas no desenvolvimento do projeto, a biblioteca **OSMnx**. Por ser baseada na biblioteca do **OpenStreetMap**, projeto de produção colaborativa de dados geo-espaciais abertos, este agrega diversos enriquecimentos e informações às vias do mapa da cidade. Sendo assim, essa ferramenta é capaz de nos fornecer representações de diversas cidades, mapeando-as como grafos, onde suas esquinas são representadas como vértices, e suas vias como arestas. Além de conter os enriquecimentos citados dentro da própria estrutura de dados criada.

4.2.3 Dados criminais

Para a cidade de São Paulo, foi possível obter os dados criminais através da plataforma SSP¹¹, responsável pela administração das polícias em todo o Estado de São Paulo no Brasil, sendo estas, polícias Civil, Militar e Técnico-Científica. O ano escolhido para obtenção dos dados foi o de 2019, último ano completo de base de dados sem exceções e interferências causadas pela pandemia durante os anos de 2020, 2021 e parte de 2022. Dentro da plataforma é possível filtrar previamente os tipos de crimes antes de baixá-los. Devido ao fato do projeto estar relacionado com pedestres e trânsito, dentre as categorias presentes na plataforma, as que mais se mostraram coerentes, foram: furto e roubo de veículos e furto e roubo de celulares, totalizando 48 arquivos, sendo um para cada mês de 2019, para cada uma das quatro categorias mencionadas. O primeiro passo para tratá-los foi fazer a limpeza dos dados vazios, crimes pertencentes às regiões metropolitanas, e colunas desnecessárias para nosso modelo, considerando como informação válida apenas linhas contendo informações do crime, data completa da ocorrência, endereço e latitude e longitude. Além disso, também foi necessário aplicar filtros para manter apenas crimes com rubricas relacionadas ao contexto de transporte público, envolvendo situações relacionadas ao trânsito, mas também aos passageiros. Com isso, ao final de todo esse processo, todos estes arquivos foram unificados em um único documento, a fim de facilitar a manipulação, com 212.079 registros criminais geolocalizados (Tabela 4.1).

Column	Description	Dtype
TIPO_CRIME	Tipo do crime	object
ANO_REGISTRO_BO	Ano de registro do boletim de ocorrência	int64
OCORRENCIA_DATA	Data da ocorrência	object
DIA_OCORRENCIA	Dia da ocorrência	int64
MES_OCORRENCIA	Mês da ocorrência	int64
OCORRENCIA_ANO	Ano da ocorrência	int64
OCORRENCIA_HORA	Hora da ocorrência	object
LATITUDE	Latitude	float64
LONGITUDE	Longitude	float64
RUBRICA	Assinatura abreviada do crime	object

Tabela 4.1: *Dataset* criminal limpo da cidade de São Paulo

Assim como em São Paulo, a ideia por trás da obtenção e limpeza do *dataset* de Curitiba foi a mesma, porém com as suas particularidades, assim como esperado. Contudo, diferente de São Paulo, os registros criminais de Curitiba estão agrupados em um único arquivo, que vem sendo incrementado desde 2009. Este encontra-se disponível na plataforma oficial de dados abertos da cidade, pertencendo à

¹¹Secretaria de Estado da Segurança Pública de São Paulo. <http://www.ssp.sp.gov.br/transparenciassp/Consulta2022.aspx>

SiGesGuarda¹². Uma das principais dificuldades encontradas ao procurar dados criminais oficiais de Curitiba foi encontrá-los com dados de geolocalização. Dentre todas as bases criminais analisadas, a informação mais próxima da localização foi o logradouro mais adjacente à ocorrência. Sendo assim, para cada crime de Curitiba, a geolocalização atribuída à ele foi obtida através da conversão do logradouro para as coordenadas mais próximas. Para fazer este mapeamento foi utilizada a ferramenta **Nominatim**¹³. Ao final do processo, foi possível enriquecer o *dataset* criminal da cidade com os dados de latitude e longitude, informações essenciais para o decorrer do projeto.

Por fim, diferente do *dataset* de São Paulo, para o qual existia um filtro para baixar os crimes de acordo com as categorias, neste *dataset* essas informações vieram todas agrupadas nesse único arquivo. Com isso, fez-se necessário selecionar e identificar manualmente as categorias dos crimes de acordo com as descrições das rubricas, filtrando os crimes descritos que tivessem alguma relação mínima com o contexto de trânsito e mobilidade de pedestres e, conseqüentemente, de passageiros. Desta forma, ao final de todo esse processo, o resultado final do *dataset* limpo conta com 12.448 registros criminais geolocalizados (Tabela 4.2).

Column	Description	Dtype
ANO_REGISTRO_BO	Ano de registro do boletim de ocorrência	float64
OCORRENCIA_ANO	Ano da ocorrência	int64
OCORRENCIA_DATA	Data da ocorrência	object
OCORRENCIA_DIA_SEMANA	Dia da semana da ocorrência	object
OCORRENCIA_HORA	Hora da ocorrência	object
OCORRENCIA_MES	Mês da ocorrência	int64
LATITUDE	Latitude	float64
LONGITUDE	Longitude	float64
APENAS_HORA	Apenas hora	int64
RUBRICA	Assinatura abreviada do crime	object
RUBRICA_SUBCATEGORIA	Subcategoria da rubrica	object

Tabela 4.2: *Dataset* criminal limpo da cidade de Curitiba

4.2.4 Dados de ônibus

Começando por São Paulo, os dados referentes às linhas de ônibus foram obtidos de forma gratuita e aberta através da plataforma **InterSCity**¹⁴. Essa base de dados conta

¹²SiGesGuarda - Sistema de Gestão da Guarda Municipal de Curitiba. Base contendo os dados das ocorrências atendidas pela Guarda Municipal de Curitiba. <https://www.curitiba.pr.gov.br/dados-abertos/busca/?termo=guarda>

¹³Esta é a interface de depuração para o mecanismo de pesquisa usado no site do OpenStreetMap. <https://nominatim.openstreetmap.org/>

¹⁴O InterSCity é um projeto de pesquisa colaborativo sediado pelo Instituto Nacional de Ciência e Tecnologia (INCT) da Internet do Futuro para Cidades Inteligentes. https://interscity.org/open_

com dois modelos de arquivos. O primeiro (Tabela 4.3) conta com quatro colunas, mas para este trabalho, apenas os campos de identificação da linha (*id*) e sequência de IDs de pontos de paradas (*stops*) serão necessários. Já o segundo (Tabela 4.4) conta com os identificadores dos pontos de paradas (*id*), e suas respectivas coordenadas (*x*, *y*), representadas em *Universal Transversa de Mercator* (UTM), sistema de coordenadas cartesianas bidimensional.

<i>id</i>	<i>interval</i>	<i>start_time</i>	<i>stops</i>
475R-10-0	1800,600,1200,3600,600,1800,720,...	22884	6905455,690006832,6902134,690006826,...
475R-10-1	1200,900,1800,0,0,900,900,1200,900,...	22734	800016536,800016589,8010019,670016568,...
...
9006-10-0	1800,1200,0,0,900,3600,1200,900,720,...	21733	110001144,110001145,110001146,110001140,...
971C-10-0	1800,1800,3600,3600,3600,600,1200,...	22775	4114446,410003415,4114444,410003413,...

2.183 linhas

Tabela 4.3: *Dataset* bruto das linhas de ônibus de São Paulo.

<i>id</i>	<i>x</i>	<i>y</i>
18848	329472.151858	7394135.6623
18849	327419.094574	7394938.93448
...
960011510	321614.454524	7388478.50362
960015841	321732.050231	7390542.39945

20.041 linhas

Tabela 4.4: *Dataset* bruto dos pontos de parada de São Paulo.

Já para a cidade de Curitiba, os dados também foram obtidos de forma gratuita, através da plataforma **URBS**¹⁵. O acesso se deu através de uma solicitação via e-mail para que pudessemos utilizar a API da plataforma, e assim, obter todas as informações de linhas de ônibus e pontos de paradas necessárias. Diferente da cidade de São Paulo, essas informações não vieram agrupadas. Com isso, apesar do *dataset* das linhas de ônibus, Tabela 4.5, vir com o identificador da linha (COD), e algumas outras informações, estas não remetem nem aos pontos de parada, nem a sequência desses. Isso se deve ao fato da API disponibilizar os dados separadamente, para cada identificador. Com isso, para cada (COD) disponível, existe um arquivo com todos os pontos de parada daquela linha, Tabela 4.6, com suas respectivas coordenadas (LAT, LON), sequência em que o ônibus passa por eles (SEQ) e outras informações sobre o ponto de parada.

Como pôde ser observado nas Tabelas 4.3, 4.4, 4.5, e 4.6, os dados brutos recebidos vieram com padrões bem distintos. Com isso, para satisfazer a ideia da construção

data/

¹⁵A Urbanização de Curitiba S/A é uma empresa de economia mista que administra equipamentos e espaços públicos, além do sistema de transporte da cidade de Curitiba. <https://www.urbs.curitiba.pr.gov.br/>

COD	NOME	SOMENTE_CARTAO	CATEG_SERVICO	NOME_COR
464	A. MUNHOZ / JD. BOTÂNICO	S	CONVENCIONAL	AMARELA
226	ABAETÉ	S	ALIMENTADOR	LARANJA
...
773	VIZINHANÇA / STA. RITA	S	ALIMENTADOR	LARANJA
657	XAXIM / CAPÃO RASO	S	ALIMENTADOR	LARANJA

288 linhas

Tabela 4.5: *Dataset* bruto das linhas de ônibus de Curitiba.

NOME	NUM	LAT	LON	SEQ	SENTIDO	TIPO	ITINERÁRIO
Rua Alcides Munhoz, 1235 - Mercês	190908	-25,4188829	-49,2995946	1	Jardim Botânico	Novo mobiliário	11133
Rua Alcides Munhoz, 1145 - Mercês	190909	-25,4192116	-49,2981128	2	Jardim Botânico	Novo mobiliário	11133
...
Rua Alcides Munhoz, 1204 - Mercês	190910	-25,4191417	-49,2981182	31	Alcides Munhoz	Placa em poste	11133
Rua Alcides Munhoz, 1235 - Mercês	190908	-25,4188829	-49,2995946	32	Alcides Munhoz	Novo mobiliário	11133

59 linhas

Tabela 4.6: *Dataset* bruto dos pontos de parada de Curitiba.

de um modelo, fez-se necessário criar uma padronização prévia dos dados, para só depois passá-los como entrada para o modelo.

Visando facilitar a manipulação das linhas de ônibus e unificação dos arquivos, o padrão escolhido para representá-las foi baseado no modelo do *dataset* de São Paulo, uma vez que este agrupava as sequências dos pontos de paradas em uma lista sequencial, para cada linha de ônibus, como mostra a Tabela 4.3.

Mesmo para o *dataset* de São Paulo, modelo base do arquivo final, foi necessário fazer um tratamento e padronização. A primeira foi em relação ao formato da geolocalização dos pontos de parada. Por estar no formato de UTM, foi preciso fazer a conversão para latitude e longitude, tanto para facilitar a manipulação, quanto para torná-la compatível com as ferramentas utilizadas no decorrer do projeto.

Já para a cidade de Curitiba, mais processos foram necessários, uma vez que cada linha de ônibus possuía um arquivo com os pontos de paradas. Neste caso, como o sistema de transporte possui 288 linhas, foi preciso baixar todos estes arquivos, ordená-los através da coluna SEQ, Tabela 4.6, para então agrupá-los em formato de lista sequencial e passá-los como coluna para o arquivo de linhas de ônibus.

Vale ressaltar que após este processo, uma validação foi necessária. Por mais que os dados brutos tenham sido rotulados como pontos de parada de São Paulo e Curitiba, após uma análise dos dados, foi possível notar linhas de ônibus e pontos de paradas intermunicipais, extrapolando o perímetro urbano considerado. Para lidar com isso, as seguintes validações foram feitas. Todos os pontos de parada foram verificados, conferindo se sua localização estava contida no perímetro da cidade ou não. A partir disso, o critério escolhido para validar as linhas, foi: se dentre os pontos de paradas presentes na lista sequencial estiver pelo menos um ponto de

parada que está fora da cidade, esta linha é descartada e rotulada como intermunicipal. Essa estratégia é importante para manter a coerência das distribuições dos pontos de parada das linhas na cidade, já que esse não possui as cidades vizinhas.

4.2.5 Dados de relevo

Para se obter os dados de relevo das cidades de São Paulo e Curitiba, a base de dados geomorfométrica do Brasil, disponibilizada publicamente pelo INPE (Instituto Nacional de Pesquisas Espaciais) em **TOPODATA**¹⁶ foi utilizada. Estes dados são importantes pois contêm as informações de relevo de todo o país. Estes são distribuídos através de quadrículas, Figura 4.1, cada uma contendo diversas linhas, que representam diferentes pontos de relevo da região selecionada. Essas linhas contam com a geolocalização do relevo, representada através de latitude e longitude, e a altitude do relevo, disponibilizada em metros, como pode ser observado na Tabela 4.7. Por isso, essas informações podem ser úteis para mapear cidades, bairros e regiões ainda menores.

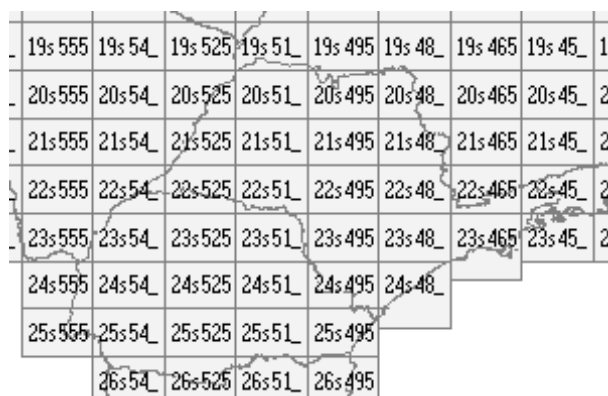


Figura 4.1: Exemplo de quadrículas no mapa do Brasil

longitude	latitude	altitude
-48,0250	-22,9750	477,0
-48,0242	-22,9750	474,0
...
-46,4767	-25,2741	0,0
-46,4758	-25,2741	0,0

8.146.800 linhas

Tabela 4.7: *Dataset* com dados brutos das altitudes (m).

Com isso, através destes dados de altitude, já se torna possível enriquecer as regiões de ambas cidades e conhecer melhor o relevo de cada uma delas. Esse

¹⁶TOPODATA - Banco de Dados Geomorfométricos do Brasil. <http://www.dsr.inpe.br/topodata/acesso.php>

processo é feito utilizando o grafo de cada uma das cidades. Desta forma, para cada vértice (esquina), a altitude mais próxima, disponível no *dataset* de altitudes, é atribuída à este vértice. Nas Figuras 4.2 e 4.3 é possível entender melhor como é o relevo de cada uma das cidades.

Altitude dos vértices do grafo da cidade de São Paulo

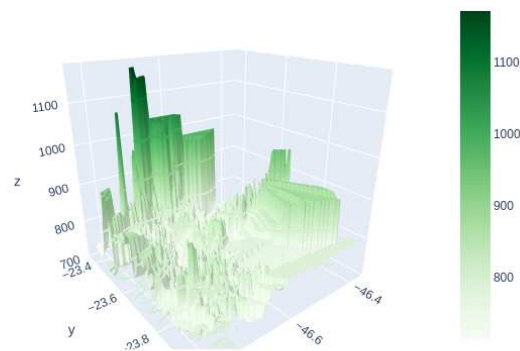


Figura 4.2: Mapa de relevo de São Paulo

Altitude dos vértices do grafo da cidade de Curitiba

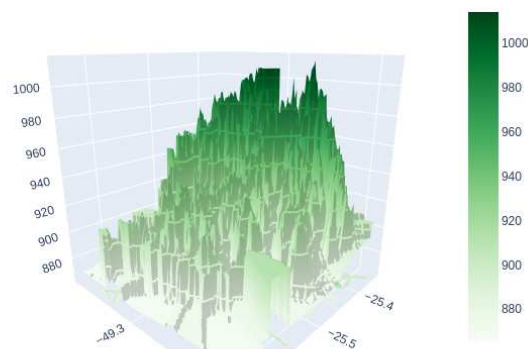


Figura 4.3: Mapa de relevo de Curitiba

4.2.6 Dados de mobilidade

Como a busca apenas dentro de sites de dados abertos não foi suficiente para obtermos acessos aos dados de mobilidade das cidades, o próximo passo foi tentar encontrar algum e-mail ou plataforma das cidades para fazer o pedido dos dados formalmente e de maneira mais específica, já informando o que queríamos e para o propósito que precisávamos. Sendo assim, na busca por portais de transparência de

São Paulo, encontramos o **e-SIC** ¹⁷. Dentro da plataforma, fizemos o pedido para o **SPTrans** ¹⁸. No conteúdo da resposta, foi explicado que para estimar o fluxo de mobilidade dos passageiros, utiliza-se a sequência de transações dos usuários obtida através do bilhete único da cidade. Isso permite mensurar a quantidade de viagens realizadas entre pontos de parada ao longo do dia. Dessa forma, para obter o fluxo, agrupam-se transações correspondentes a uma viagem completa, incluindo eventuais transferências. Posteriormente, nos foi enviado um link de uma pasta no Google Drive contendo dados de viagens do dia 21 de junho de 2022, que representa um dia útil típico na cidade (terça-feira), juntamente com o mapeamento da rede de transportes que operou neste dia.

Na Tabela 4.8 é possível perceber em detalhes o arquivo com as informações de mobilidade recebida. Nota-se que para cada par de ponto de parada de partida (*stop_id_inicio*) e de destino (*stop_id_fim*), existem várias colunas que representam as horas do dia, de 0 a 23 horas. Com isso, cada coluna possui o número de passageiros que fizeram aquele trajeto no intervalo de tempo equivalente.

<i>stop_id_inicio</i>	<i>stop_id_fim</i>	0	1	2	3	...	18	19	20	21	22	23
18848	18849	NaN	NaN	NaN	NaN	...	22.0	22.0	10.0	4.0	2.0	NaN
18848	18850	NaN	NaN	NaN	NaN	...	32.0	20.0	4.0	6.0	4.0	NaN
...
960011510	960011473	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
960011510	960011475	NaN	2.0	NaN	NaN	...	NaN	2.0	NaN	NaN	NaN	NaN

1.363.005 linhas

Tabela 4.8: *Dataset* com dados de fluxos de passageiros entre pares de pontos de parada.

Entretanto, para o propósito deste projeto, o nível de granularidade em horas não é necessário, pois o uso da mobilidade será utilizado apenas representando o fluxo diário. Desta forma, as colunas de 0 a 23 apenas foram agrupadas para uma única coluna, contendo a soma de todo o fluxo distribuído nas mesmas (Tabela 4.9).

<i>stop_id_inicio</i>	<i>stop_id_fim</i>	<i>total_dia</i>
18848	18849	184.0
18848	18850	214.0
...
960011510	960011473	2.0
960011510	960011475	4.0

1.362.262 linhas

Tabela 4.9: *Dataset* simplificado com coluna de total de fluxo do dia.

¹⁷e-SIC – Sistema Eletrônico do Serviço de Informação ao Cidadão da cidade de São Paulo. O objetivo do e-SIC é facilitar o exercício do direito fundamental de acesso às informações públicas. <https://esic.prefeitura.sp.gov.br/>

¹⁸SPTrans - São Paulo Transportes S/A. Órgão responsável pela gestão do sistema de transporte público na cidade de São Paulo. <https://www.sptrans.com.br/sptrans/>

Para complementar as informações neste primeiro *dataset*, foi disponibilizado um segundo arquivo (Tabela 4.10), responsável por mapear os IDs dos pontos de parada com coordenadas e informações reais da cidade. Este arquivo é importante, pois para integrar essas informações da cidade com o grafo utilizado no projeto, as colunas de coordenadas geográficas (latitude e longitude) são essenciais.

stop_id	stop_name	stop_desc	stop_lat	stop_lon
18848	Clínicas	NaN	-23,554022	-46,671108
18849	Vila Madalena	NaN	-23,546498	-46,691141
...
960011509	Av. Dr. Guilherme Dumont Villares, 50	Ref.: R Alfredo Mendes Da Silva/ Av Professor ...	-23,604681	-46,747021
960011510	R. Alfredo Mendes Da Silva, 499	Ref.: R Cedro/ Av Doutor Guilherme Dumont Vill...	-23,604420	-46,748357

21.034 linhas

Tabela 4.10: Informações complementares do campo *stop_id*.

Assim como na cidade de São Paulo, também procuramos pela plataforma da cidade de Curitiba. No próprio site da cidade é possível entrar na seção de lei de acesso à informação e solicitá-la. Nessa plataforma oficial **Serviço Digital de Curitiba** ¹⁹ é possível fazer um pedido à algum órgão específico, assim como em São Paulo. Com isso, fizemos o pedido à unidade responsável pelo transporte público da cidade, **URBS** ²⁰. A resposta se deu pela própria plataforma, entretanto, as informações desta já eram conhecidas e já havíamos acessado as recomendações sugeridas anteriormente. Porém, nessa requisição nos passaram o e-mail de contato direto com a empresa ²¹. Com isso, fizemos a mesma solicitação, desta vez, via e-mail, e obtivemos sucesso. Posteriormente, recebemos um arquivo CSV, conforme mostrado na Tabela 4.11, com as informações do dia 8 de março de 2023 (quarta-feira), dia útil típico na cidade.

ID_Cartao_cripto	DATA_COMPLETA	NR_LATITUDE	NR_LONGITUDE	...
1	2023-03-08 11:55:21	-25,558555	-49,329323	...
2	2023-03-08 11:56:41	-25,55956	-49,331719	...
..
33106	2023-03-08 12:29:22	-25,498913	-49,286031	...
252955	2023-03-08 12:32:01	-25,461446	-49,274255	...

441.447 linhas x 9 colunas

Tabela 4.11: *Dataset* de mobilidade da cidade de Curitiba.

Devido à falta de informação de destino onde o passageiro desce, tendo apenas o ponto de partida, uma forma de inferir seu deslocamento é tentar encontrar algum outro momento do dia que este mesmo usuário parte de um novo destino.

¹⁹Serviço Digital de Curitiba - Acesso à Informação. <https://servicodigital.curitiba.pr.gov.br/>

²⁰URBS - Urbanização de Curitiba S/A. <https://www.urbs.curitiba.pr.gov.br/transporte/red-e-integrada-de-transporte>

²¹E-mail de contato com a central de relacionamento da empresa URBS. relacionamento@urbs.curitiba.pr.gov.br

Entretanto, vários obstáculos são encontrados nesse processo, pois, segundo o funcionário da URBS que nos auxiliou, é comum uma pessoa passar o cartão para outra, situação que contabiliza duas vezes a mesma partida para um único passageiro. Entretanto, a próxima partida daquele mesmo usuário seria do mesmo lugar, gerando um deslocamento errado. Outro caso que pode induzir à inferência ao erro são deslocamentos rápidos, que uma mesma pessoa vai em um intervalo de tempo curto em vários lugares, seja para resolver problemas particulares ou à trabalho. Desta forma, algumas recomendações nos foram passadas. A primeira é restringir o intervalo das duas viagens deste passageiro em pelo menos 4 horas. Com isso, seria mais garantido que este não estaria nos casos citados anteriormente. E outra validação necessária está relacionada com a quantidade de vezes que este usuário usa o transporte no dia. Para o padrão de mobilidade que estamos interessados, é relevante deslocamentos semelhantes à saída de casa do passageiro para o trabalho, e o retorno, após um intervalo de tempo maior. Ou seja, para um mesmo passageiro, ter apenas dois registros de uso do transporte durante o dia, em um intervalo maior ou igual à 4 horas. O resultado desta inferência pode ser visto na Tabela 4.12. Agora, com ambos *datasets* preparados, já se torna possível aplicar o fluxo de mobilidade na solução RACIONAL desenvolvida neste trabalho.

stop_id_inicio	stop_id_fim	total_dia
5471847535	4828352575	191
4828352575	4828352575	140
...
847468170	269204919	1
8582770084	1362783978	1

126.500 linhas

Tabela 4.12: *Dataset* de mobilidade da cidade de Curitiba enriquecido com inferência.

4.3 Construção de Abstrações

4.3.1 Clusterização dos dados criminais

O algoritmo DBSCAN foi utilizado para mapear geograficamente as áreas criminais distribuídas pela cidade. Esta escolha ocorreu devido à habilidade do algoritmo de gerar agrupamentos (*clusters*) com diferentes formatos e tamanhos, e por ter bom resultado de clusterização mesmo na presença de ruídos Santos et al. (2017). Além disso, esta solução também possui a vantagem de gerar a clusterização preservando a localização original dos crimes, característica essencial para o contexto das linhas de ônibus e pontos de paradas, pois garante que aquela região, dita insegura, está

realmente na localização analisada, permitindo construir soluções reais para cada região avaliada.

Com isso, após esta etapa de seleção e limpeza de crimes coerentes para o contexto analisado, detalhados na Subseção 3.2.2, basta utilizá-los em conjunto com o algoritmo DBSCAN, para gerar e demarcar os *clusters* criminais pela cidade.

O algoritmo do DBSCAN requer dois parâmetros para seu funcionamento: ϵ (raio entre vizinhos) e v (quantidade mínima vizinhos). A escolha destes valores, inicialmente, foram $v = 36$ e $\epsilon = 100\text{m}$, que foram obtidos empiricamente ao analisar diferentes combinações para o *dataset* criminal do ano de 2014 da cidade de São Paulo, com aproximadamente 523 mil crimes, selecionados de acordo com o contexto de transporte público. Estes valores já foram utilizados em Almeida et al. (2022), proporcionando uma distribuição de *clusters* que, além de mapear bem locais inseguros, permite espaço para se encontrar caminhos mais seguros alternativos. Entretanto, para a base de dados utilizada neste trabalho, referente ao ano de 2019, os crimes selecionados somam um total de aproximadamente 212 mil ocorrências, ou seja, menos da metade dos que foram utilizados no trabalho de referência. Com isso, ao aplicar os mesmos parâmetros, apresentados acima, para a cidade de São Paulo, os *clusters* gerados não corresponderam às configurações e distribuições esperadas. Sendo assim, foi necessário encontrar as melhores combinações de parâmetros novamente. Ao final deste processo, os valores que mais se aproximaram do esperado, foram $v = 14$ e $\epsilon = 85\text{m}$, resultando em 2028 *clusters* criminais distribuídos pela cidade.

Entretanto, com a clusterização criada, ainda não é possível representá-la no mapa da cidade, nem prosseguir com o desenvolvimento do RACIONAL, pois ainda não possui o formato compatível de visualização e utilização necessária, que neste caso, equivale à polígonos. Desta forma, mais um passo é aplicado para que seja possível obtê-las. A conversão utilizada para gerar os polígonos é conhecida como **Convex Hull** ²², que basicamente consiste em pegar um conjunto de pontos clusterizados, e através do menor número possível de pontos externos, gerar um polígono convexo que contenha todos os outros pontos dentro. Com isso, ao finalizar este passo, o número de polígonos disponíveis passa a ser 1961, o que será usado de fato no desenvolvimento do trabalho.

Na Figura 4.4a é possível observar o resultado da clusterização. Contudo, vale ressaltar que o resultado obtido possui alguns pontos a serem analisados. Se compararmos os dois maiores *clusters* gerados na Figura 4.4a com os bairros da cidade de São Paulo (Figura 4.4b) podemos observar que só o *cluster* central da cidade, sozinho, é maior que vários bairros adjacentes juntos. O outro *cluster*,

²²Convex Hull: A convex hull is the smallest convex polygon containing all the given points. <https://www.geeksforgeeks.org/convex-hull-using-divide-and-conquer-algorithm/>

localizado à esquerda desse maior, também possui um tamanho considerável, quase cobrindo um bairro sozinho. Com isso, diante dessa situação, optamos por remover estes dois *clusters* maiores. O motivo da remoção é que posteriormente, na etapa de realocação, se mantidos, resultariam em casos onde não seria possível realocar nenhum ponto de parada dentro de todos esses bairros que estão contidos dentro deles. Além de semanticamente, representar que em toda essa área, nenhum local seria seguro para trafegar. Entretanto, também traz alguns pontos que devem ser mencionados. Dentro das métricas, a região central da cidade é afetada, pois grande parte dos crimes foram agrupados nestes *clusters* removidos, com isso, pode acontecer casos onde alguns bairros da região central não possuem *clusters* criminais, pois como estavam dentro do que foi removido, não restou mais nenhum na região. O resultado final da remoção destes dois *clusters* desbalanceados pode ser visto na Figura 4.4c.

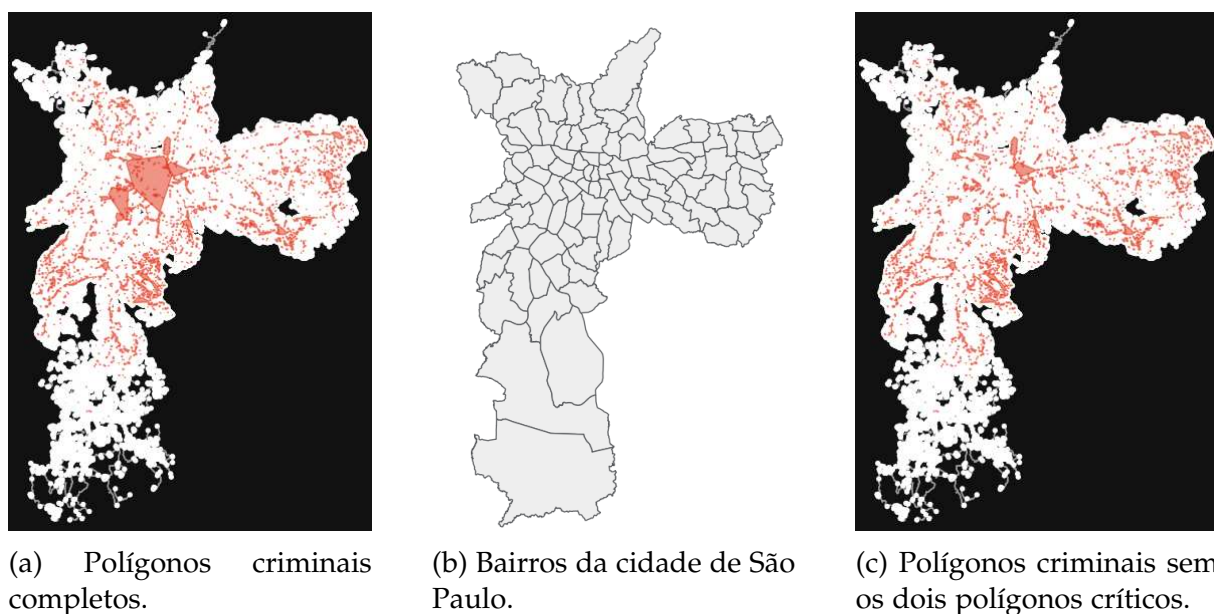


Figura 4.4: Análise dos polígonos criminais gerados na cidade de São Paulo.

Assim como foi necessário descobrir os parâmetros empiricamente para São Paulo, para Curitiba também não foi diferente. Entretanto, dois obstáculos foram encontrados no processo. O primeiro foi a quantidade desproporcional de crimes entre as duas cidades, assim como foi já detalhado anteriormente. O segundo foi não possuir uma referência da mesma cidade, assim como em São Paulo, que mesmo não sendo possível aproveitar os parâmetros anteriores, foi possível replicar uma distribuição parecida, que já havia sido testada em Almeida et al. (2022). Desta forma, o critério utilizado para encontrar a melhor combinação de parâmetros foi testar diversos parâmetros e analisar aqueles que geravam distribuições de *clusters* de forma equilibrada pela cidade, assim como São Paulo. Além disso, dentre esses selecionados, também foi feita uma análise da média das suas áreas, com o objetivo



(a) Polígonos criminais.



(b) Bairros da cidade de Curitiba.

Figura 4.5: Análise dos polígonos criminais gerados na cidade de Curitiba.

de filtrar os que mais se aproximavam de São Paulo. Com isso, após validar esses dois critérios, os parâmetros escolhidos para Curitiba foram: $v = 3$ e $\varepsilon = 250\text{m}$. Como resultado, foram gerados 360 *clusters*, quantidade bem inferior à de São Paulo. Além disso, ao aplicar o algoritmo de Convex Hull, o número de polígonos gerados caiu consideravelmente, resultando em 136 polígonos. O resultado final pode ser visto na Figura 4.5a, assim como a configuração dos bairros da cidade (Figura 4.5b). Percebe-se que, visualmente, o maior polígono mantido é semelhante aos que foram removidos na cidade de São Paulo. Entretanto, o critério de remoção foi baseado em um valor máximo de área, para ambas cidades, e neste caso, este polígono está situado dentro do intervalo permitido, e por isso está mantido no resultado final.

Para complementar o entendimento relacionado aos polígonos gerados pela cidade, foram gerados resultados complementares aos apresentados nas Figura 4.4 e Figura 4.5. Apesar de ser possível notar superficialmente a distribuição dos polígonos, a magnitude de cada região se torna quase imperceptível, visto que o processo de projetar visualmente os polígonos nos bairros é impreciso. Logo, para contornar essa questão, foi gerada uma análise da distribuição destes polígonos dentro de cada bairro, através de um mapa de calor. Com isso, é possível entender melhor em quais regiões estão concentrados os maiores números de polígonos criminais.

Na cidade de São Paulo (Figura 4.6) é importante lembrar que apesar de grande parte estar mais situada nas regiões periféricas, grande parte dos crimes da região central foram convertidos nos dois polígonos maiores (outliers), que precisaram ser removidos. Em quantidade representam apenas dois polígonos, mas conceitualmente,

continham muitos crimes. A remoção destes, acabou influenciando na ocorrência de dois bairros centrais ficarem sem nenhum polígono. Já para Curitiba (Figura 4.7) a maior parte dos polígonos ficaram distribuídos na região leste da cidade, entretanto, o bairro com maior quantidade de polígonos está localizado no sentido oposto, na região oeste.

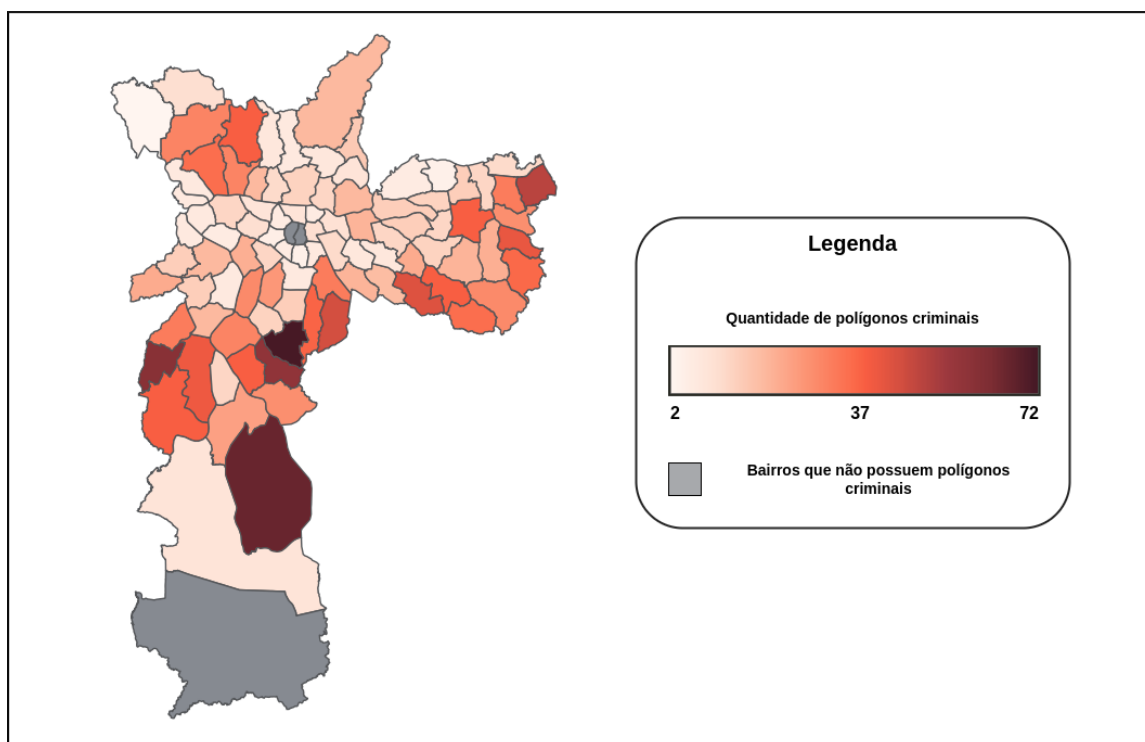


Figura 4.6: Mapa de calor dos bairros que mais tiveram polígonos criminais na cidade de São Paulo.

4.3.2 Construção do grafo enriquecido da cidade

O enriquecimento do grafo foi feito utilizando as funcionalidades disponíveis na ferramenta OSMnx. Para simplificar a modelagem do problema, dados como relevo e pontos de parada precisaram ser mapeados nos nós das esquinas mais próximas, para se adequar melhor à representação do grafo utilizada pela ferramenta OSMnx.

Com isso, após o enriquecimento mencionado, as demais manipulações no grafo foram realizadas para otimizar o processo de desenvolvimento. Assim, por usufruir de uma estrutura de dados complexa e rica em informações sobre cada esquina e via das cidades, toda manipulação e identificação de quaisquer informações se tornava restrita à criação de uma instância para cada cidade, independentemente se seriam utilizadas as operações disponíveis desse grafo ou se apenas uma simples recuperação de informação seria realizada. Com isso, além da estrutura de dados cara, o processo de recuperação de informações dos vértices e arestas também se

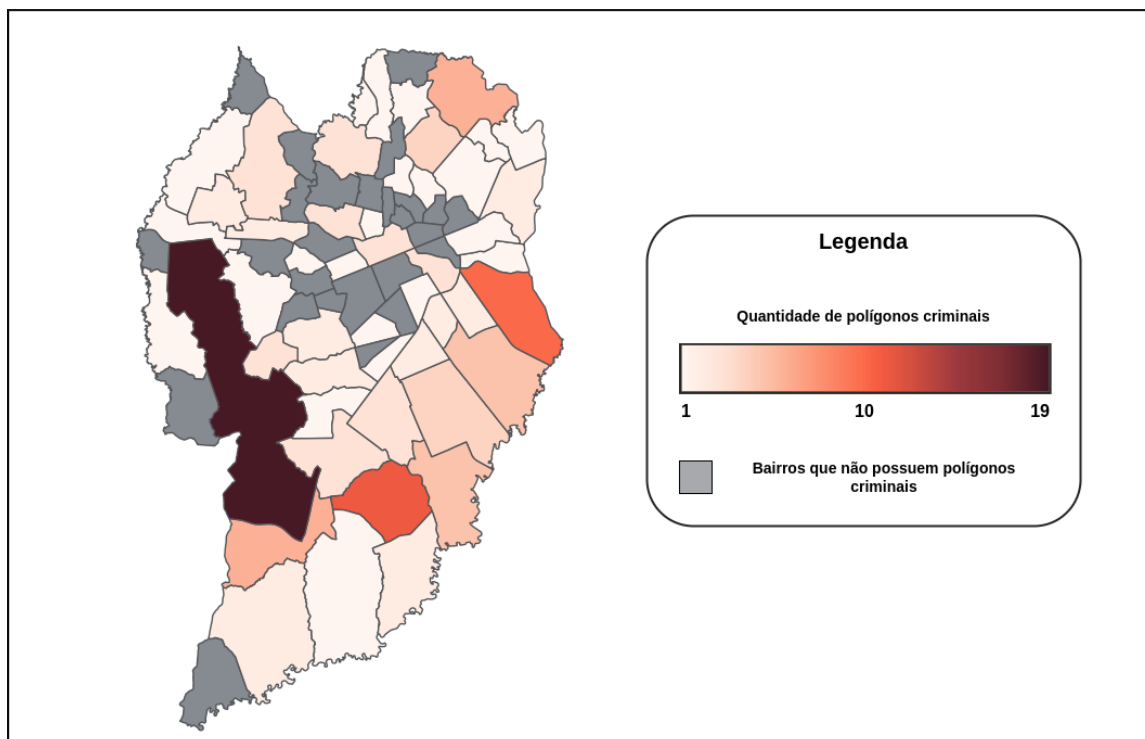


Figura 4.7: Mapa de calor dos bairros que mais tiveram polígonos criminais na cidade de Curitiba.

tornava trabalhoso e demorado. Logo, optamos por criar um *dataset* que continha essas mesmas informações, e outras extras, como a altitude de cada vértice do grafo, por exemplo. Entretanto, visando torná-lo mais acessível tanto para a instância quanto para a busca de informações específicas, a ideia é reservar o uso do grafo apenas quando necessário, especialmente em situações que demandem suas operações e algoritmos.

Outro ponto observado foi em relação à identificação de pontos de paradas localizados dentro de polígonos criminais. Este passo exige a instanciação do grafo e a necessidade de aplicação de diferentes cálculos sobre ele. Porém, uma vez realizados, dado que os polígonos são gerados apenas uma vez para as duas cidades, essa informação sempre seria a mesma, independente do momento em que fosse consultada, pois os vértices do grafo não perderiam suas localizações e identificadores durante o processo. Assim sendo, ao relacionar esse problema com o citado anteriormente, optou-se por criar colunas extras no *dataset* de representação dos grafos, adicionando esses cálculos custosos, para fazê-los apenas uma vez e depois apenas carregá-los para consulta, proporcionando um ganho significativo de tempo no processo de implementação.

Desta forma, a ideia foi aplicada à cada vértice dos grafos das duas cidades, sempre guardando para cada configuração de v e ϵ os cinco polígonos mais próximos, já ordenados pela menor distância, e a informação se o vértice analisado

estava contido em cada um desses polígonos ou não. Com isso, com todas essas informações disponíveis no *dataset*, e de fácil acesso e recuperação, o ganho obtido no tempo de implementar e dar manutenções no modelo diminuiu drasticamente. O resultado final deste enriquecimento pode ser observado na Tabela 4.13.

vertice	cidade	lat	long	ponto_parada	ref_alt	altitude	poligonos_mais_proximos
573641	São Paulo	-23,5629	-46,7039	False	41,80	727,0	{'raio_85_vizinhos_14': [{'id_poligono': 972, ...}]}
573643	São Paulo	-23,5674	-46,7022	True	27,89	728,0	{'raio_85_vizinhos_14': [{'id_poligono': 1602, ...}]}
...
1007...	Curitiba	-25,4027	-49,2460	False	37,40	937,0	{'raio_250_vizinhos_3': [{'id_poligono': 1995, ...}]}
1008...	Curitiba	-25,3903	-49,2581	False	37,17	953,0	{'raio_250_vizinhos_3': [{'id_poligono': 2023, ...}]}

156.252 linhas

Tabela 4.13: *Dataset* final com todas informações.

Por fim, além dessas operações realizadas diretamente no grafo, também foi necessário implementar uma outra melhoria para mitigar outro gargalo encontrado no processo de criação das rotas seguras realocadas. Ao se deparar com uma situação de realocação de ponto de parada, é necessário checar nas redondezas deste se existe algum outro ponto que atenda as condições necessárias para poder substituí-lo. Entretanto, realizar essa operação no grafo é muito custosa computacionalmente. Contudo, apesar de ser inevitável no processo, deixá-la em tempo de execução, ou seja, durante a criação da rota, se torna uma abordagem ruim, pois à cada modificação ou correção, todo este cálculo seria realizado novamente, atrasando muito no processo de desenvolvimento.

Desta forma, a ideia utilizada para resolver este problema também foi criar um *dataset* onde essas informações fossem calculadas apenas uma vez, para estarem sempre disponíveis rapidamente em qualquer etapa do desenvolvimento. Dado que salvar essa informação uma única vez é coerente, pois conceitualmente, os pontos de paradas não perdem suas referências geoespaciais, assim como os vértices do grafo também não. Contudo, por ser uma etapa antes do desenvolvimento em si, não é possível saber quais mapeamentos serão requisitados, fazendo com que seja necessário criar quaisquer tipos de possibilidade para garantir a cobertura de qualquer situação futura. Apesar desse processo ser muito caro, por ser feito apenas uma única vez, torna-se mais viável do que deixá-lo em tempo de execução. Na Tabela 4.14 é possível observar o mapeamento criado. Como houve a necessidade de combinar todas as possibilidades de pares de vértices, ou seja, para cada vértice, fazer o mapeamento com todos os outros existentes, o resultado final culminou em um *dataset* com mais de 2 milhões de pares de vértices, incluindo vértices das duas cidades.

ponto_parada_origem	ponto_parada_destino	dist_linha_reta_m	dist_trajeto_m	cidade
573643	25618442	1188,47	3928,68	São Paulo
25618442	573643	1188,47	1229,49	São Paulo
...
10046431381	10046431382	57,27	57,65	Curitiba
10046431382	10046431381	57,27	57,78	Curitiba

2185912 linhas

Tabela 4.14: *Dataset* final com todas informações.

4.4 Construção das Rotas Seguras

4.4.1 Construção de todas as rotas possíveis com pontos de parada originais

Antes de explicar em detalhes o processo de construção de rotas seguras, é relevante destacar que, embora tenhamos apresentado a manipulação dos dados obtidos nos estágios anteriores de obtenção e limpeza, foram necessárias adaptações para integrar esses dados à estrutura de grafos fornecida pela biblioteca OSMnx.

Começando pela quantidade de linhas de ônibus disponíveis, São Paulo conta com 2069 linhas, e Curitiba com 676 linhas. Entretanto, ao aplicá-las na ferramenta citada, algumas não puderam ser convertidas corretamente para a estrutura de grafos disponibilizada, pois devido à aproximações dos pontos de paradas para as esquinas mais próximas, algumas acabaram não ficando coerentes, sendo necessário removê-las. Além disso, mesmo possuindo rotas que não tiveram nenhum tipo de impedimento relacionado à isso, ainda sim tiveram que ser descartadas por serem linhas intermunicipais, o que seria um problema em relação à falta de dados disponíveis para as outras cidades. Entretanto, o número de linhas removidas foi pequeno, restando 1987 linhas para São Paulo, com cada uma tendo em média 42 pontos de parada, e 629 linhas para Curitiba, contando com cerca de 27 pontos de paradas, em média.

Com isso, já com as linhas de ônibus prontas para serem aplicadas no modelo, a estrutura criada para representá-las pode ser vista na Tabela 4.15.

ID	ROTAS_CONVERTIDAS_ID_GRAFO
475R-10-0	[1734285568, 1734285531, 1734285575, 173428565...
2013-10-0	[1804029276, 4692263357, 1367924287, 180402949...
...	...
9006-10-0	[1804087083, 4758828263, 4758828303, 180408753...
971C-10-0	[4782091398, 4782088980, 4782091369, 478209136...

126.500 linhas

Tabela 4.15: Linhas de ônibus prontas para o modelo.

A partir desta representação, já é possível aplicar cada linha ao modelo, permitindo a construção dos diferentes cenários. Um diferencial desta

implementação utilizada é que essa se beneficia da necessidade de passar pelos pontos de paradas das linhas de par em par. Com isso, à cada iteração, são calculadas todas as opções de trechos possíveis, e em seguida, são ranqueados de acordo com seu índice K . Sendo assim, ao final de cada iteração, é armazenado apenas o que é de interesse para a avaliação, como por exemplo, a **ROTA MAIS SEGURA** e a **ROTA MAIS CURTA**, cujo índice K e a distância são, respectivamente, os menores possíveis.

Para as **ROTAS MAIS SEGURAS REALOCADAS** o processo também acontece de forma semelhante, tendo apenas um passo anterior, responsável por avaliar os pontos de paradas e realocá-los ainda em nível de linha de ônibus. Ou seja, o modelo desconhece a existência de pontos realocados, pois apenas interpreta linhas de ônibus modificadas, mas que seguem a mesma estrutura apresentada em ???. Com isso, ao final do processo, cada linha obteve sua respectiva rota, de acordo com os cenários pré-estabelecidos. Por fim, com o intuito de facilitar a análise posterior dos resultados, foi construído um *dataset* único, com todas as linhas das duas cidades agrupadas.

4.4.2 Construção de todas as rotas possíveis utilizando pontos de parada alternativos

Seguindo a proposta do modelo, no que tange a implementação das condições de realocação dos pontos de parada, o primeiro requisito é definir o *Raio* e *Distância* máxima permitidas para as duas cidades, de acordo com a Subseção 3.4.3.

Para as cidades de São Paulo e Curitiba, foi definido um raio máximo de $Raio = 250m$, representando a distância limite em linha reta entre os pontos de parada avaliados. Já para a distância percorrida pela via, foi definida como $Distancia = 1000$ metros, visto que o deslocamento real possui uma distância maior do que aquela em linha reta.

Vale ressaltar que a escolha desses valores foi feita de forma empírica, avaliando a distribuição dos pontos de paradas pela ferramenta de grafo, e complementando com base em situações reais de distância, trânsito e transportes públicos locais.

Além disso, também como critério de avaliação de realocação de pontos de paradas existe a questão da impedância, que também foi definida empiricamente. Contudo, diferente dos valores obtidos anteriormente, que eram facilmente explicáveis e interpretáveis, a obtenção desse se mostrou mais complexa.

Desta forma, para se obter este valor da impedância, o processo se passou através de caminhadas em lugares com aclives acentuados. Para isso, foi utilizado o aplicativo **Strava**²³. Nele é possível registrar o deslocamento e ganho de elevação do

²³Strava: O Strava conecta milhões de corredores, ciclistas, caminhantes e outras pessoas ativas por meio dos esportes que elas adoram. <https://www.strava.com/?hl=pt-BR>

percurso. Logo, foi possível obter o ângulo de inclinação dos testes feitos através destes registros. Dentre os testes empíricos, o escolhido para se obter a angulação foi o de 180 metros de deslocamento e 9 metros de ganho de elevação, registrado em uma via movimentada e frequentada por diferentes tipos de pessoas e idades. Para se obter o ângulo deste caso de teste, foi criada uma aproximação com um triângulo retângulo, onde a hipotenusa foi 180 e o cateto oposto 9. Com isso, o ângulo obtido foi de aproximadamente $2,87^\circ$. Desta forma, utilizando o percurso completo, ou seja, percorrendo os 180 metros no trajeto inclinado, a impedância obtida através da Equação 3.2, foi de aproximadamente 1477. Sendo assim, a partir deste valor, o valor de referência passou a ser um arredondamento para 1500. Desse modo, agora com o valor de referência, os valores dos ângulos e das distâncias para os diferentes cenários possíveis podem variar, desde que fiquem abaixo de 1500.

Para que fosse possível gerar os cálculos de impedância em quaisquer regiões das cidades, o primeiro passo foi enriquecer cada vértice das cidades com a altitude disponível mais perto. Neste momento, o *dataset* apresentado em ?? foi utilizado. Por esses dados de relevo não estarem necessariamente nos mesmos lugares que os vértices estão, foi necessário aplicar a abordagem de identificação das referências de altitude mais próximas de cada vértice, para então atribuí-las a ele.

No entanto, durante o desenvolvimento do trabalho Almeida et al. (2022), que naquele momento, estava avaliando a viabilidade de criar rotas seguras para ônibus públicos em São Paulo, por meio de uma adaptação do trabalho Ladeira et al. (2019), foi possível notar que alguns processos se mostraram muito caros computacionalmente, e identificar os pontos de paradas candidatos é um deles. Com isso, durante este processo, encontrar os pontos candidatos em tempo de construção de rota é custoso computacionalmente e problemático, pois qualquer configuração alterada e que requisitasse uma nova execução exigiria com que todo o processo fosse repetido novamente. Deste modo, posteriormente, isso acabou se tornando um empecilho muito grande, sendo necessário levar essa execução para um *Cluster UFV*²⁴ e deixá-la dias executando o algoritmo de construção de rotas para se obter o resultado.

Sendo assim, nesta nova etapa de construção do modelo, para contornar os problemas apresentados, foi realizado um mapeamento de todos os pares possíveis de pontos de parada de cada cidade. O objetivo foi calcular diversas informações de uma única vez e deixá-las disponíveis como uma espécie de gabarito para sempre que um par fosse solicitado. Com essa nova abordagem, foi possível executar essa aplicação localmente, utilizando o notebook apresentado na Seção 4.1, em um curto intervalo de tempo (poucas horas). Além disso, pudemos usufruir dos benefícios já

²⁴Cluster UFV: Máquinas dedicadas à disponibilização de recursos computacionais para pesquisas <https://dct.ufv.br/uso-do-ambiente/>

mencionados, como fácil depuração e reexecução do processo de criação de rotas seguras, caso necessário.

Dentre as informações agregadas, os dados de relevos puderam ser usados para identificar os ângulos e impedâncias. Para realizar este cálculo, primeiro foi necessário encontrar a diferença de altura entre dois pontos de paradas, e em seguida, a distância em linha reta (euclidiana). Com isso, foi possível aplicar o mesmo passo mencionado anteriormente, e descobrir o ângulo entre os dois pontos de paradas analisados. Por fim, bastou aplicar a Equação 3.2 sobre o ângulo obtido e a distância entre os dois pontos de paradas. Entretanto, a distância em linha reta entre dois pontos no contexto de cidades não representa a realidade, pois a maneira que os bairros e ruas são distribuídos não permitem o acesso em linha reta de forma fácil, sendo possível apenas quando estão em uma mesma rua ou esquinas subsequentes. Desta forma, por se tratar de linhas de ônibus, para que pares de pontos de paradas se encontrem nessa situação é muito pouco recorrente. Com isso, dentre as informações de deslocamento disponíveis, a opção mais coerente com a realidade foi utilizar a distância do trajeto mais curto como referência.

Logo, com os valores já definidos, aplicar o algoritmo 2 foi suficiente para avaliar todas as sugestões de pontos de paradas para realocação e selecionar apenas aqueles que respeitassem os limites pré-estabelecidos. Com isso, as linhas de ônibus puderam ser modificadas para então serem enviadas para a construção das rotas, assim como já mencionado na subseção anterior (Seção 4.4.1).

4.5 Avaliação

4.5.1 Filtragem das rotas construídas

Nesta etapa do estudo de caso, para tornar o processo mais eficiente devido à estratégia implementada, todas as rotas foram geradas em paralelo com base nas linhas de ônibus das duas cidades. Dessa forma, a seleção e filtragem das rotas, com base na constante K , não foram realizadas apenas ao final da iteração por todos os pontos de parada das linhas, mas sim à medida que os pares de pontos de parada de cada linha eram percorridos.

À medida que esses pares eram iterados, todas as opções de rotas eram geradas e recebiam seus respectivos índices de criminalidade. O mesmo processo era aplicado ao cenário de realocação, onde, antes de construir a rota, eram feitas as realocações de pontos, quando possível. Posteriormente, o filtro era aplicado, selecionando apenas o trecho com o menor índice criminal.

Por fim, é relevante destacar que, para as **ROTAS MAIS CURTAS**, o processo também seguiu a mesma abordagem, construindo-as durante a iteração dos pares

de pontos de parada das linhas, mas desta vez, priorizando as rotas com menores comprimentos em metros.

No contexto de trechos com maiores fluxos de mobilidade, o objetivo também foi aproveitar as estruturas já criadas, inclusive todos os cenários existentes. Com isso, para obtê-los, foi necessário iterar todos os pontos de parada das linhas, par a par, e verificar qual deles possuía o maior índice de mobilidade, da origem (ponto 1) ao destino (ponto 2), com base nas bases de dados de viagens das duas cidades. Ao identificá-los, foi necessário apenas salvá-los com o identificador da linha e a rota entre os pares correspondente, sendo necessário apenas ajustar os resultados com base no novo recorte.

4.5.2 Cálculos de métricas por linha

Conforme apresentado na Seção 3.5, devido à baixa variação do principal indicador de segurança K , outras estratégias foram utilizadas. Dentre elas, uma manipulação geral dos dados possibilitou a criação de métricas relativas que contribuíram no processo de análise e interpretação dos resultados. Essa manipulação está relacionada com o uso de um cenário como base; neste caso, o de **ROTAS MAIS CURTAS** é utilizado como a situação atual das cidades avaliadas. Desta forma, os resultados médios deste cenário para cada rota são fixados, e os demais, das rotas seguras, são gerados em relação a este. Assim, ao final do processo, é possível avaliar em que medida cada métrica apresentou melhora ou piora em relação ao cenário base.

4.5.3 Cálculo de métricas por trecho

Para as métricas por trecho, a abordagem foi semelhante, sendo necessário apenas representá-las de forma que o algoritmo não percebesse que estava manipulando um trecho em vez de uma rota completa. Para isso, bastou tratar todos os trechos gerados como rotas contendo apenas dois pontos de parada. Por fim, é importante ressaltar que, devido a essa característica dos trechos de possuir apenas dois pontos de parada, algumas métricas que dependem de mais pontos não podem ser calculadas.

Capítulo 5

Resultados

5.1 Avaliando as rotas completas

Neste capítulo serão discutidos os principais resultados obtidos ao aplicar o modelo RACIONAL desenvolvido e explicado anteriormente nas cidades de São Paulo e Curitiba. Dessa maneira, todas as métricas e resultados avaliados a seguir estarão sempre relacionados a essas duas cidades, acompanhadas das principais observações e peculiaridades de cada uma delas.

Para entender os resultados relacionados às linhas de ônibus, é importante conhecer melhor a distribuição e magnitude dos dados tratados pelo modelo, além dos principais impactos das alterações sugeridas, para somente depois comparar os resultados dos cenários estabelecidos. Começando pelo nível de segurança dos pontos de parada, tem-se que na cidade de São Paulo, 27,8% dos 42 pontos de parada (média por rota) estão localizados dentro de *clusters* criminais. Já em Curitiba, apenas 6,5% dos 27 pontos de parada (média por rota), estão. Com isso, ao aplicar a solução proposta, espera-se que um número maior de pontos de parada seja realocado para a cidade de São Paulo, pois proporcionalmente, está mais sujeita à situações de realocações, além de contar com um número maior de pontos de paradas.

Por se tratar de uma solução que visa melhorar a segurança de um modelo de transporte já existente, entendemos que a melhor forma de apresentar os resultados seria comparando o quanto cada métrica foi melhor ou pior do que o cenário de rotas mais curtas, utilizado para representar a realidade atual. Com isso, ao invés de apenas apresentar os gráficos de cada cenário lado a lado e apenas comparar o valor absoluto de cada, optamos por fazer uma análise relativa e ver, de fato, o quanto cada métrica foi melhor proporcionalmente do que o cenário atual, de modo a facilitar a análise.

Sendo assim, para conseguir criar este cenário relativo, foi necessário pegar cada linha de ônibus, isolar o cenário mais curto, utilizado para representar a realidade da cidade, e então calcular os demais cenários em relação à este. Desta forma, cada cenário fica com a diferença positiva ou negativa em relação ao modelo atual. Com isso, ao plotar os resultados, todos os ganhos e perdas podem ser identificados de

forma mais precisa e explícita.

Apesar da impedância ser um requisito para construir as rotas visando um esforço equilibrado para os passageiros, vale ressaltar o impacto real que as rotas receberam por ela. Para isso, foi criada a métrica **Impedância média dos pontos de parada realocados**. Embora o valor máximo permitido tenha sido definido como 1500, na prática, os resultados de impedância foram bem menores, sendo em média 390 para São Paulo, quase 4x menor que o limite, e 148 para Curitiba, mais de 10x menor (Figura 5.1). Vale ressaltar que os resultados dos cenários mais curto e mais seguro são iguais pois representam o contexto sem esforço extra para os passageiros, uma vez que os pontos de paradas são os mesmos para ambos.

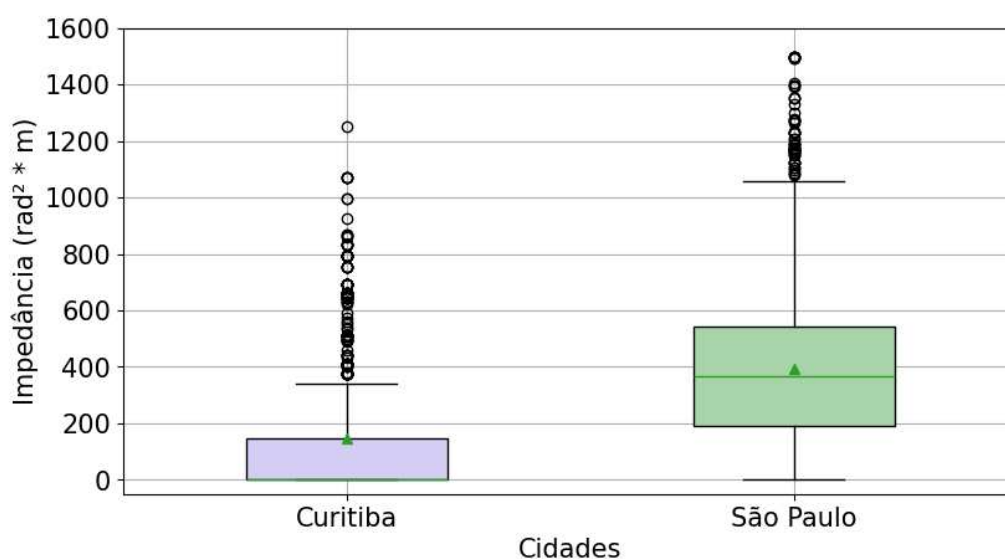


Figura 5.1: Acréscimo de impedância nas sugestões de realocação.

Esses resultados são relevantes para mostrar que mesmo adicionando essa camada a mais de impedância, com o intuito de garantir a experiência do passageiro parecida com a realidade, ou seja, sem adicionar muito esforço extra para se deslocar até o novo ponto de parada, as outras abordagens adotadas na primeira etapa deste trabalho foram eficientes para garantir a qualidade na identificação e realocação de pontos candidatos, uma vez que estiveram bem abaixo do limite máximo adicionado.

Com isso, apesar de gerar alternativas de **ROTAS MAIS SEGURAS**, considerando pontos dentro do limite de impedância, ainda é necessário observar que se além da baixa impedância apresentada, essas modificações foram de fato significativas e representativas. Para isso, foi feita uma análise comparativa de quantos **Pontos de paradas foram removidos de regiões criminais de cada linha de ônibus**. Como já mencionando anteriormente, essas modificações são aplicadas apenas em rotas realocadas, logo, para esta métrica, também são esperadas mudanças apenas nesse cenário. Desta forma, na Figura 5.2 é possível observar que em média, quase 20% do total de pontos de paradas das linhas puderam ser

realocados. Em números absolutos, isso quer dizer que, em média, 7 dos 42 pontos de paradas foram realocados para regiões mais seguras em cada linha analisada. Para Curitiba, também houveram ganhos. Na Figura 5.2 aproximadamente 4% dos pontos de parada foram realocados de regiões identificadas como inseguras, representando 0.8 dos 27 pontos, em média.

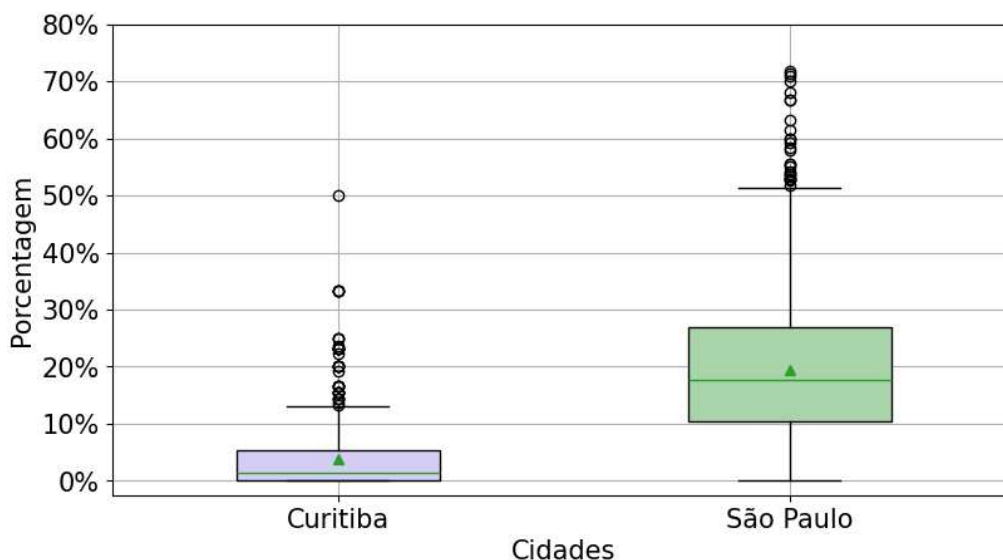


Figura 5.2: Pontos de paradas removidos de dentro de *clusters* criminais.

Nota-se que houve uma diferença significativa nos resultados de ambas cidades. Isso se deve ao fato da discrepância encontrada nas quantidades de crimes e *clusters* distribuídos pela cidade. Como São Paulo possui uma quantidade muito maior, cerca de 27.8% dos pontos de paradas das linhas originais estão localizados dentro de regiões identificadas como inseguras, enquanto apenas 6.5% de Curitiba estão.

Apesar do principal parâmetro para definir e classificar as rotas entre mais ou menos seguras ser a variável K , o valor absoluto da mesma não apresenta diferenças notórias quando calculada para diferentes rotas, pois as variações estão em casas decimais distantes. Desta forma, surgiu a necessidade de apresentar essas diferenças de forma mais fácil e visível, utilizando parâmetros mais voltados ao contexto de trânsito.

Sendo assim, uma das formas de representar estes ganhos obtidos pela variável k foi construindo a métrica de **Soma das distâncias sequenciais dentro dos mesmos *clusters***. Essa métrica é calculada com base nas distâncias percorridas por pares de vértices adjacentes de uma mesma rota que se encontram dentro dos mesmos *clusters* criminais. Dessa forma, para cada linha, todos os cenários de rotas são submetidos a essas somas das distâncias sequenciais. Ao final do processo, é obtida a média dessas distâncias para cada cenário avaliado, considerando todas as linhas. Dada às limitações apresentadas na ferramenta utilizada, esta foi a maneira mais próxima de

simular deslocamento real dentro de regiões identificadas como inseguras. Com isso, começando por São Paulo, os resultados apresentados na Figura 5.3 nos mostram que as duas rotas que visam maior segurança conseguiram evitar caminhos contidos dentro de *clusters* criminais, percorrendo uma distância menor em relação à rota mais curta, por isso o resultado negativo. Para o cenário de rota mais segura, sem realocação, o ganho médio foi de 84 metros, ou seja, deixou de percorrer essa distância dentro de regiões identificadas como inseguras. Já para o cenário de rota mais segura realocada, o ganho foi ainda maior, totalizando 334 metros.

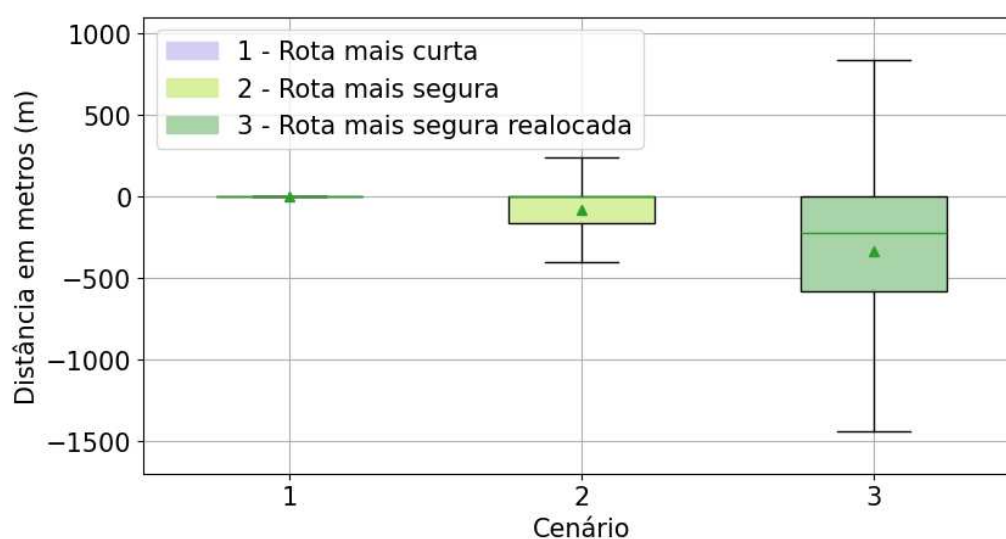


Figura 5.3: Soma das distâncias sequenciais dentro dos mesmos *clusters* de São Paulo.

Para Curitiba, os resultados também seguiram o mesmo comportamento apresentado na cidade de São Paulo, Figura 5.4. Para o cenário de rota mais segura, o ganho médio foi de 14 metros, já para o cenário de rota mais segura realocada, a média foi de 52 metros. Esta diferença significativa pode ser explicada pela variação de comprimento das rotas das duas cidades, próxima métrica a ser apresentada, e pela diferença significativa de *clusters* distribuídos pela cidade, o que impacta na menor quantidade de vértices contidos dentro de regiões identificadas como inseguras. Para a cidade de São Paulo, cerca de 28% dos vértices das rotas estão contidos dentro de *clusters*, já Curitiba, conta com apenas 6,4%.

Outra métrica importante para avaliar a solução RACIONAL é o **Acréscimo total das rotas** para cada cenário. Uma vez que o fator com maior prioridade dos cenários de rota mais segura e rota mais segura realocada é a segurança dos passageiros através da variável K , evitar regiões que possuem um grau de periculosidade maior se torna a primeira opção no lugar de comprimento da rota. Com isso, é esperado que esses cenários apresentem ganhos em relação à rota original, definida como mais curta. Na Figura 5.5 é possível observar que o cenário rota mais segura apresenta um

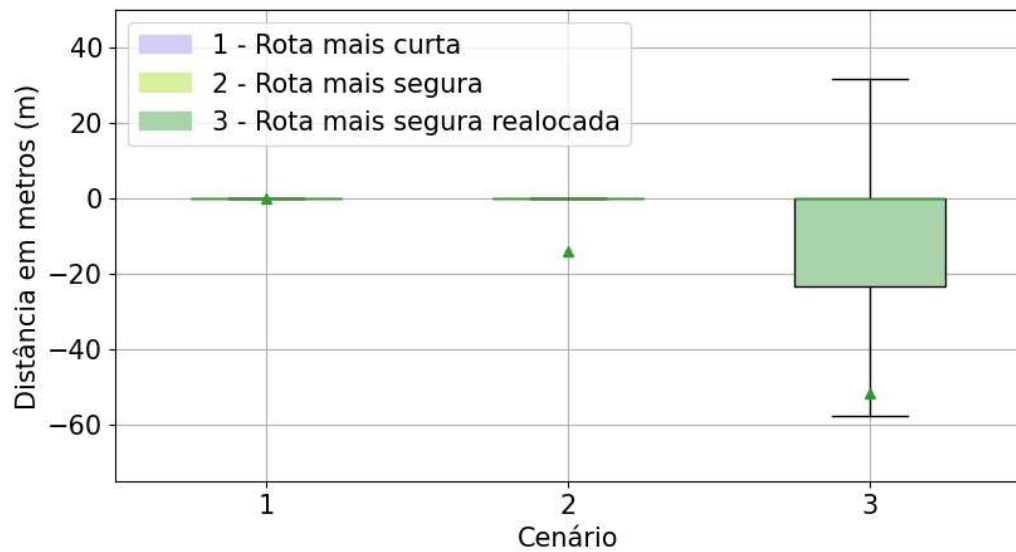


Figura 5.4: Soma das distâncias sequenciais dentro dos mesmos *clusters* de Curitiba.

acréscimo médio de 1.125 metros, já o cenário mais segura realocada possui uma média maior, de 2.164 metros, este ainda conta com o deslocamento dos pontos de paradas para outros elegíveis, fator extra que adiciona ainda mais deslocamento ao percurso, além dos desvios presentes nas rotas dos trechos mais seguros. À título de comparação, o cenário rota mais curta, usada como referência, conta com 19.480 metros de comprimento médio, desta forma, os acréscimos percentuais dos dois cenários foram equivalentes à 5,78% para rota mais segura, e 11,11% para a rota mais segura realocada.

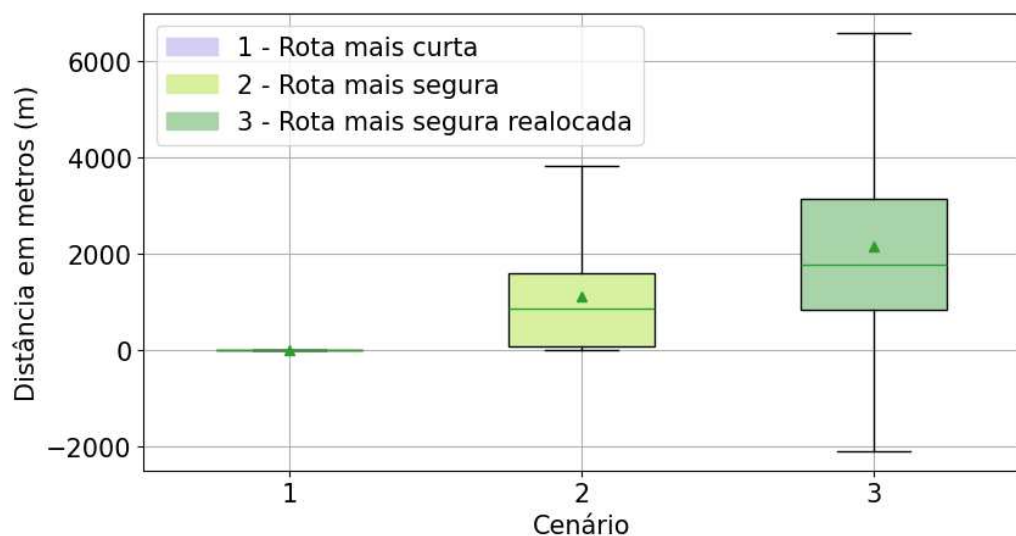


Figura 5.5: Acréscimo total das rotas na cidade de São Paulo.

Para Curitiba os resultados Figura 5.6 também apresentaram um acréscimo em

relação ao cenário base de rotas mais curtas, entretanto, com uma diferença menor em relação aos cenários mais seguro e mais seguro realocado. Percebe-se que a menor ocorrência de pontos de paradas localizados dentro de *clusters* diminuiu a necessidade de mudanças dos pontos originais, e conseqüentemente, refletiu em uma melhor qualidade nos pontos de paradas elegíveis, beneficiando o algoritmo e mitigando o acréscimo em deslocamento entre os dois cenários mais seguros, ao contrário de São Paulo. Desta forma, o acréscimo médio para o cenário rota mais segura foi de 597 metros, enquanto o da rota mais segura realocada foi de 593 metros. Se comparado ao comprimento total do cenário base, que conta com 13.357 metros, os acréscimos foram de 4,48% para o cenário mais seguro, e 4,45% para o cenário mais seguro realocado.

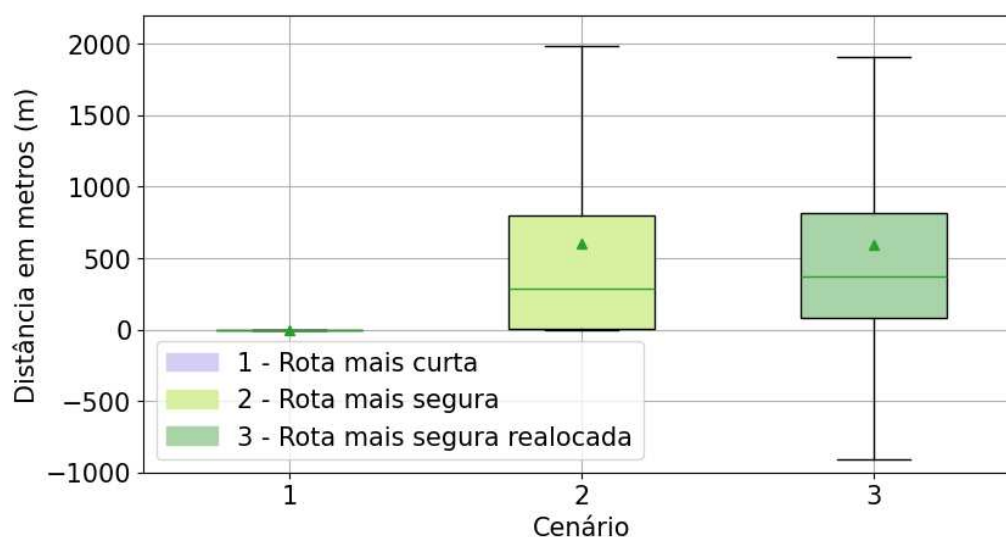


Figura 5.6: Acréscimo total das rotas na cidade de Curitiba.

Para entender a distribuição desses deslocamentos de forma mais precisa, foi criada uma outra métrica, que ao invés de avaliar o comprimento da rota como um todo, desmembra esta com base nas distâncias dos seus trechos, registrando todos esses deslocamentos entre pares de pontos de parada da rota. A partir desses valores, é feita a média para descobrir o **Acréscimo médio dos trechos de cada rota**. Sendo assim, com esse valor é possível construir a métrica de média de comprimento dos trechos da rota. Na Figura 5.7 é possível observar que o acréscimo médio por trecho da rota em São Paulo, é baixo, sendo 28,18 metros para o cenário mais seguro, e 76,41 metros para o cenário mais curto realocado. Em termos percentuais, o acréscimo por trecho é aproximadamente 5,73% para a rota mais segura, e 15,54% para a rota mais segura realocada, uma vez que o comprimento médio do cenário mais curto é de aproximadamente 491,84m. Em termos práticos, o acréscimo de tempo para o passageiro não seria muito significativo, pois de acordo com o

Mobilize Brasil ¹, um ônibus na cidade de São Paulo em horário de pico consegue se locomover na média de 16km/h. Com isso, no pior caso, envolvendo o cenário em que um passageiro embarca no primeiro ponto de parada e desembarca apenas no último, o acréscimo de tempo para a rota mais segura realocada completa seria de aproximadamente 11 minutos.

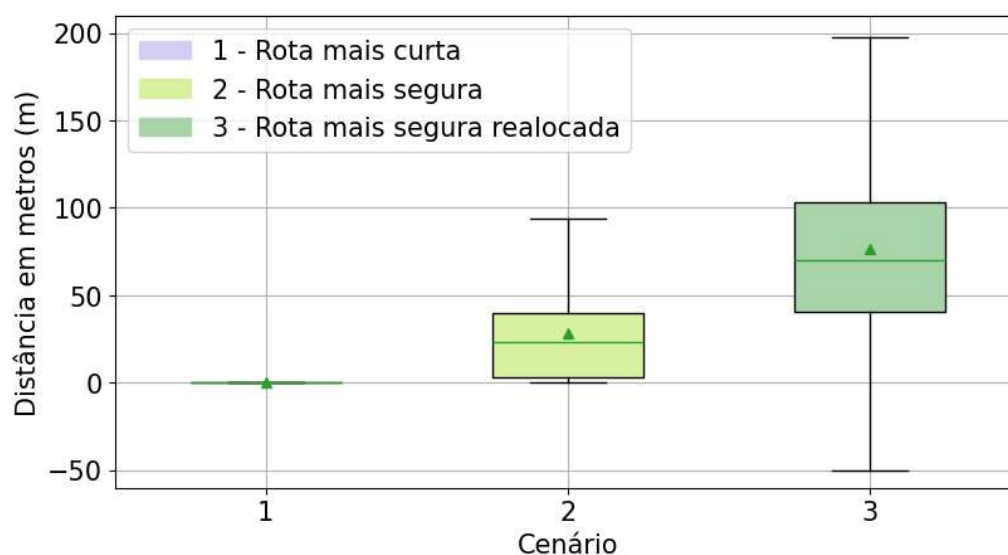


Figura 5.7: Média dos acréscimos dos trechos de São Paulo.

Já em Curitiba, os resultados Figura 5.8 também foram semelhantes aos apresentados na métrica de comprimento total, também apresentando um acréscimo médio muito próximo para os dois cenários, sendo 41,09 metros (5,49%) para as **ROTAS MAIS SEGURAS**, e 41,4 metros (5,53%) para as **ROTAS MAIS SEGURAS REALOCADAS**, tendo em média 748,46m para a rota mais curta. Em termos práticos, assim como o apresentado na cidade de São Paulo, o acréscimo de tempo para o passageiro seria ainda menor, pois de acordo com o Estadão Summit Mobilidade ², a velocidade média de ônibus na cidade de Curitiba é de 26,7km/h. Com isso, ao simular o pior caso, considerando o uso do serviço para a linha completa, o tempo de acréscimo seria de aproximadamente 2,5 minutos, resultado ainda menor.

Apesar do acréscimo de comprimento apresentado na métrica anterior, além das melhorias já apresentadas, a solução também é capaz de distanciar as esquinas que o ônibus percorre dos centros de *clusters* mais próximos. Na Figura 5.9 é possível

¹Mobilize Brasil Org - Primeiro portal brasileiro inteiramente dedicado ao tema da mobilidade urbana sustentável. <https://www.mobilize.org.br/noticias/11963/velocidade-media-dos-onibus-fica-estagnada-em-16-kmh-na-cidade-de-sp.html>

²Estadão Summit Mobilidade - O maior e mais relevante evento sobre o futuro da mobilidade no Brasil. <https://summitmobilidade.estadao.com.br/compartilhando-o-caminho/onibus-de-curitiba-como-a-cidade-popularizou-o-brt/>

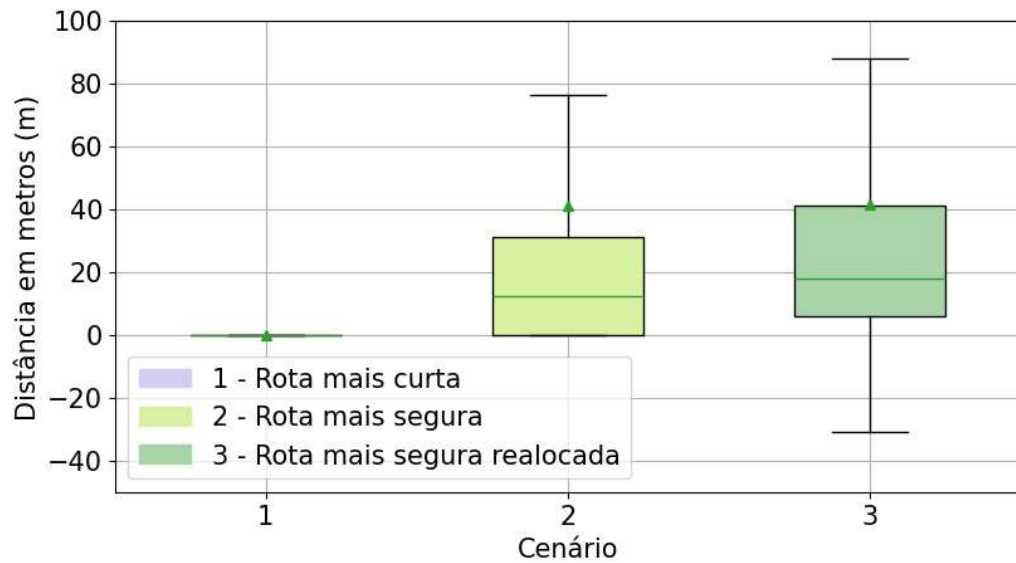


Figura 5.8: Média dos acréscimos dos trechos de Curitiba.

perceber através da métrica **Média das distâncias dos nós em relação aos centros de clusters** o ganho de distância para os dois cenários na cidade de São Paulo, sendo 4,26 metros para as **ROTAS MAIS SEGURAS** e 6,88 metros para as **ROTAS MAIS SEGURAS REALOCADAS**. Esse resultado ressalta que além das sugestões de realocação dos pontos de paradas, o trajeto, como um todo, também está mais seguro, não restringindo a solução à apenas o conceito de realocação do cenário 3.

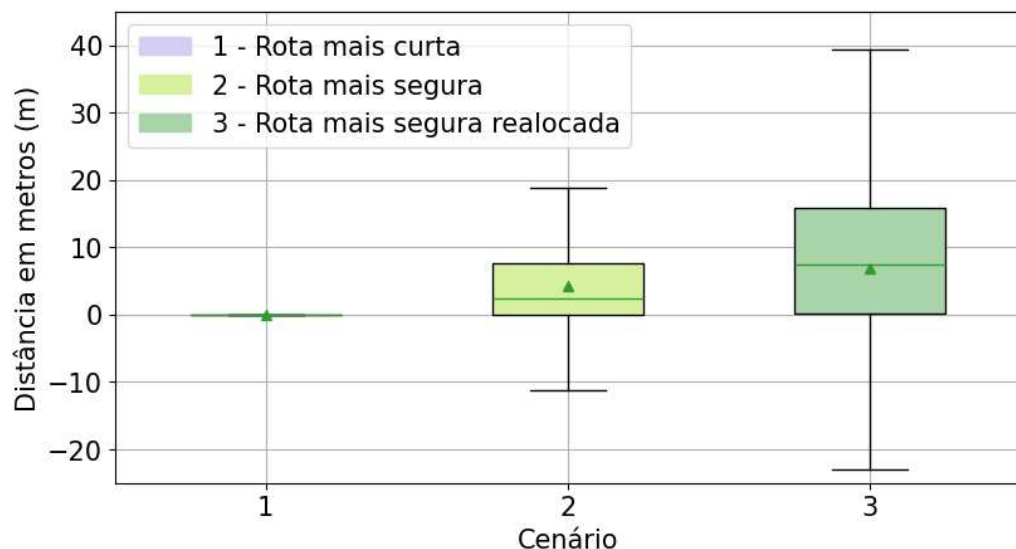


Figura 5.9: Média das distâncias dos nós em relação aos centros dos *clusters* de São Paulo.

Os resultados para Curitiba também apresentaram um ganho semelhante, entretanto, com médias maiores. Para as **ROTAS MAIS SEGURAS**, houve um

distanciamento médio de aproximadamente 9 metros, já para as **ROTAS MAIS SEGURAS REALOCADAS**, o ganho foi de 12 metros.

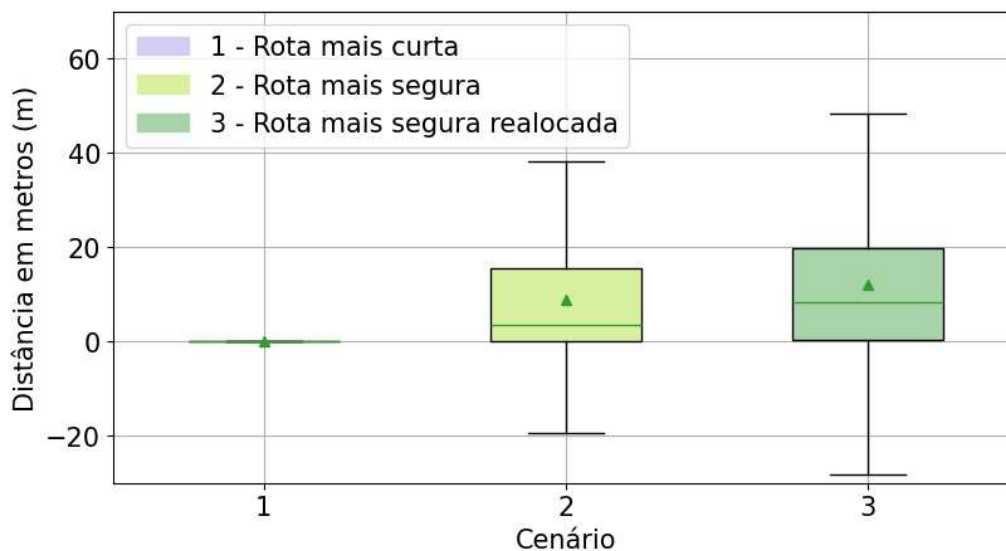


Figura 5.10: Média das distância dos nós em relação aos centros dos *clusters* de Curitiba.

Para avaliar a qualidade do cenário de realocação, também foi criada a mesma métrica exibida anteriormente, porém, considerando apenas as esquinas que representam pontos de paradas nas rotas. Essa métrica é denominada **Média das distâncias dos pontos de paradas em relação aos centros dos clusters**. Para a cidade de São Paulo, o ganho médio de distanciamento dos pontos de paradas dos centros dos *clusters* no cenário de realocação foi de 18,35 metros, como pode ser observado na Figura 5.11.

Já para Curitiba, o ganho foi menor, totalizando 2,55 metros em média, Figura 5.12. Esse resultado é coerente, pois para a cidade de Curitiba, a média de pontos de paradas das linhas localizados dentro de *clusters* é de 6,5%, já em São Paulo, são aproximadamente 27,8%. Além disso, o número de pontos de parada que foram realocados e removidos de dentro de *clusters* é bem menor, totalizando 0,8 (3,71%), justamente por não haver a mesma necessidade de realocação como na cidade de São Paulo, que demandou um número de trocas muito maior, tendo margem para realocar cerca de 7,19 (20%) pontos.

5.2 Cálculo de métricas por trecho

Além dos resultados que já envolvem as linhas de ônibus completas, foi criada uma análise mais voltada para identificar as partes mais relevantes do trajeto, ou seja, regiões de cada rota que mais poderiam impactar os passageiros durante o percurso.

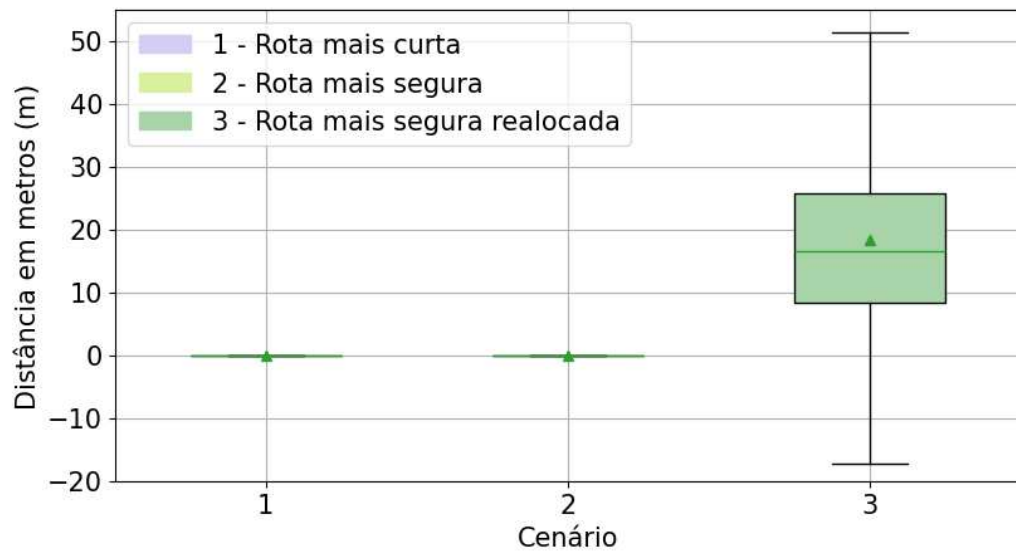


Figura 5.11: Média das distâncias dos pontos de parada em relação aos centros dos *clusters* de São Paulo.

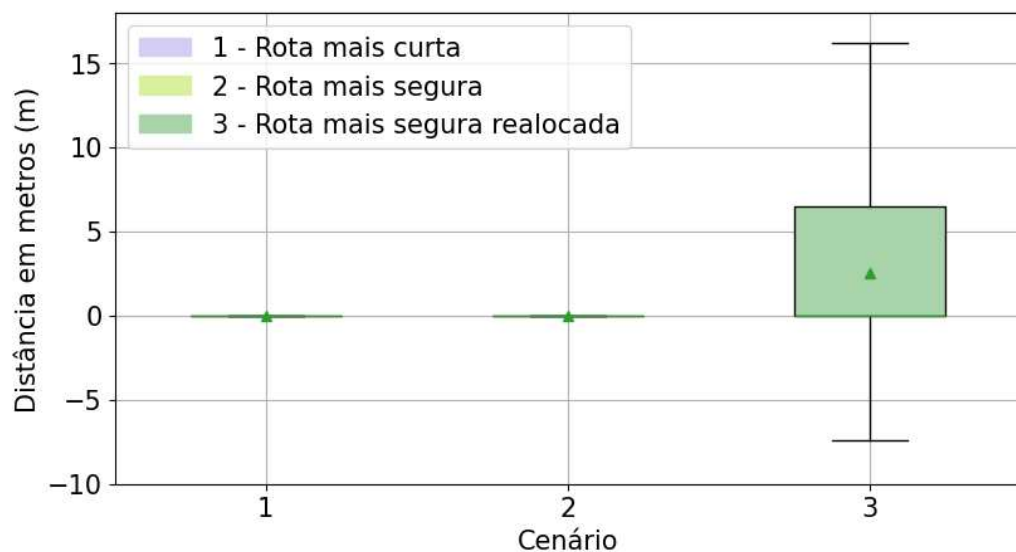


Figura 5.12: Média das distância dos pontos de parada em relação aos centros dos *clusters* de Curitiba.

Para isso, essas regiões foram definidas como os trajetos que mais recebem fluxos de pessoas. Para desenvolver essa análise, foi necessário mapear em cada rota criada a quantidade de pessoas que se locomoveram entre pares de pontos de paradas, através de trechos, de modo a construir um resultado apto para ser ranqueado. Isso pôde ser feito utilizando as informações de mobilidade, já apresentadas no decorrer do trabalho, Subseção 4.2.6.

Sendo assim, para cada rota foi feita uma lista de fluxos de pessoas, contendo a quantidade de pessoas que se locomoveram entre cada trecho durante o dia. Em

seguida, o ranqueamento é construído através da ordenação dessa mesma lista. Desta forma, para essa análise, como o foco foi voltado para os principais trechos, a escolha do item ordenado foi destinada sempre para o trecho de maior soma de mobilidade, representando sempre o top 1 de fluxos de pessoas da rota.

Com os trechos identificados, a próxima estratégia adotada foi a criação de um *dataset* independente e compatível com o modelo RACIONAL desenvolvido no projeto. Isso é necessário e importante pois garante a compatibilidade do processo de construção das rotas e dos resultados. Logo, ao final do processo, os trechos são entendidos pelo modelo como novas rotas, contendo sempre dois pontos de parada e as demais informações até essa etapa, também construídas de forma equivalente às originais.

Uma vez pronta, essa estrutura permite gerar resultados da mesma forma, porém com as ressalvas esperadas, como por exemplo não ser possível construir a métrica de média dos comprimentos dos trechos da rota completa, pois a rota possui sempre apenas um trecho, o do top 1 de mobilidade.

Com isso, os resultados a seguir serão fundamentados nas mesmas métricas previamente apresentadas, mas sempre satisfazendo a condição de trecho de maior mobilidade, assim como explicado anteriormente. Vale ressaltar, mais uma vez, que o intuito desses resultados são validar se os trechos mais importantes das linhas também se comportam de acordo com os padrões observados nos resultados ou não, ou seja, avaliar a viabilidade e impacto mais significativo na região da rota que mais recebe fluxos de pessoas.

O resultado da média da soma das distâncias sequenciais que o ônibus percorre dentro dos mesmos *clusters* apresenta um comportamento bem semelhante ao já apresentado para as linhas completas. Isso nos mostra que mesmo para as regiões que têm um fluxo de pedestres maior, podendo ser uma região de maior relevância tanto para a cidade, quanto para rota, é possível trazer um ganho no nível de segurança nos dois cenários de maior segurança avaliados. Na Figura 5.13 é possível notar que a média de ambos cenários se encontra negativa, representando justamente que para a **ROTA MAIS SEGURA**, o ônibus deixou de andar 5,98 metros dentro de *clusters* criminais. Já para as **ROTAS MAIS SEGURAS REALOCADAS**, o ganho foi maior, sendo de 16,77 metros, em média.

Para Curitiba os resultados também foram de melhoria, permitindo um menor deslocamento dentro de *clusters* criminais para os dois cenários. Para as **ROTAS MAIS SEGURAS**, o ganho foi de 0,5 metros, já pra as **ROTAS MAIS SEGURAS REALOCADAS**, o resultado foi maior, totalizando 13,02 metros, Figura 5.14.

Para oferecer esse ganho de segurança, parte do comprimento total da rota é influenciado, pois as soluções mais seguras acabam recebendo um acréscimo no processo de distanciamento dos *clusters* criminais. Desta forma, assim como nos

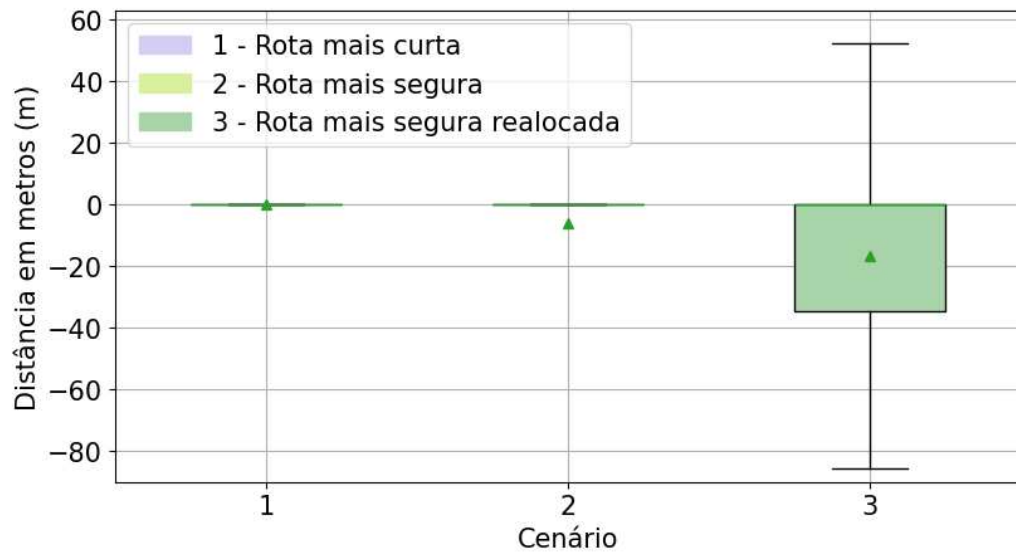


Figura 5.13: Soma das distâncias sequenciais dentro dos mesmos *clusters* criminais para trechos com maiores fluxos de pessoas de São Paulo.

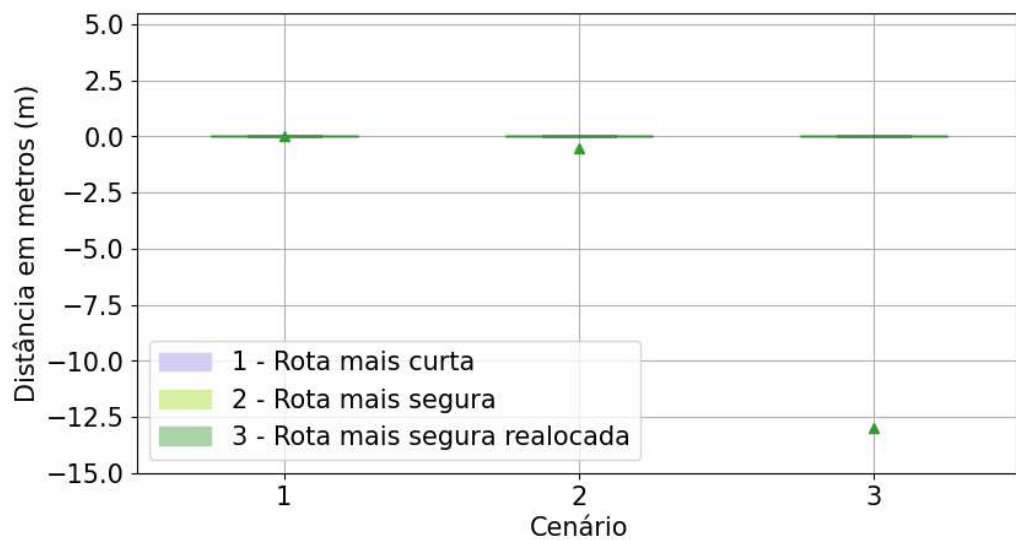


Figura 5.14: Soma das distâncias sequenciais dentro dos mesmos *clusters* criminais para trechos com maiores fluxos de pessoas de Curitiba.

resultados de linhas completas, para o contexto do trecho com maior fluxo de mobilidade, o ganho em comprimento também aparece. Na Figura 5.15 nota-se o acréscimo para a cidade de São Paulo, porém, para ambos cenários, são deslocamentos pequenos, adicionando 30,26 metros (4%) para as **ROTAS MAIS SEGURAS**, e 91 metros (12,14%) para as **ROTAS MAIS SEGURAS REALOCADAS**.

Para Curitiba, Figura 5.16, é possível observar o mesmo comportamento de maior similaridade dos cenários de **ROTAS MAIS SEGURAS**, apresentando menor diferença entre si, se comparados aos resultados da cidade de São Paulo. Em média são 45,6 metros (5,74%) para as **ROTAS MAIS SEGURAS**, e 45,66 metros (5,39%)

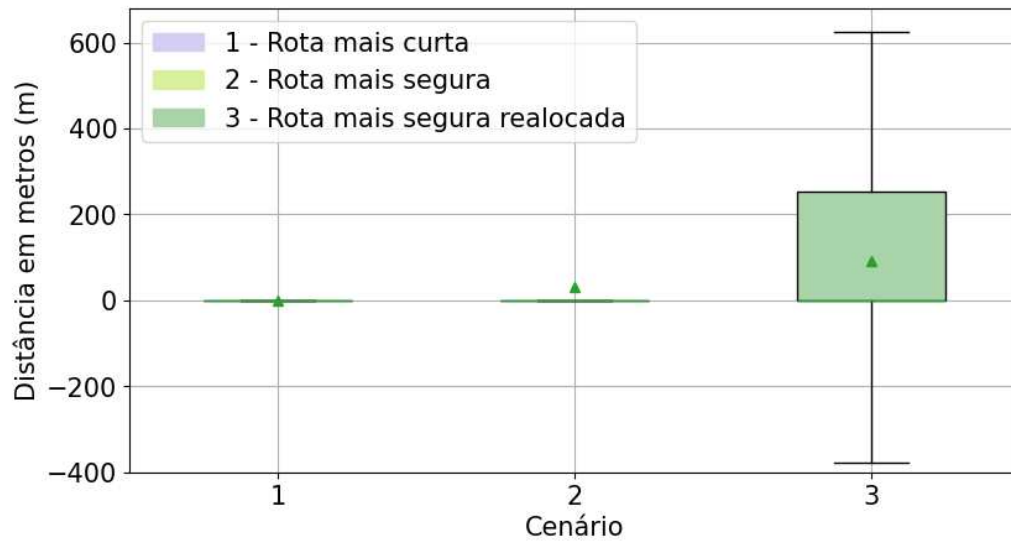


Figura 5.15: Acréscimo total do trecho na cidade de São Paulo.

para as **ROTAS MAIS SEGURAS REALOCADAS**.

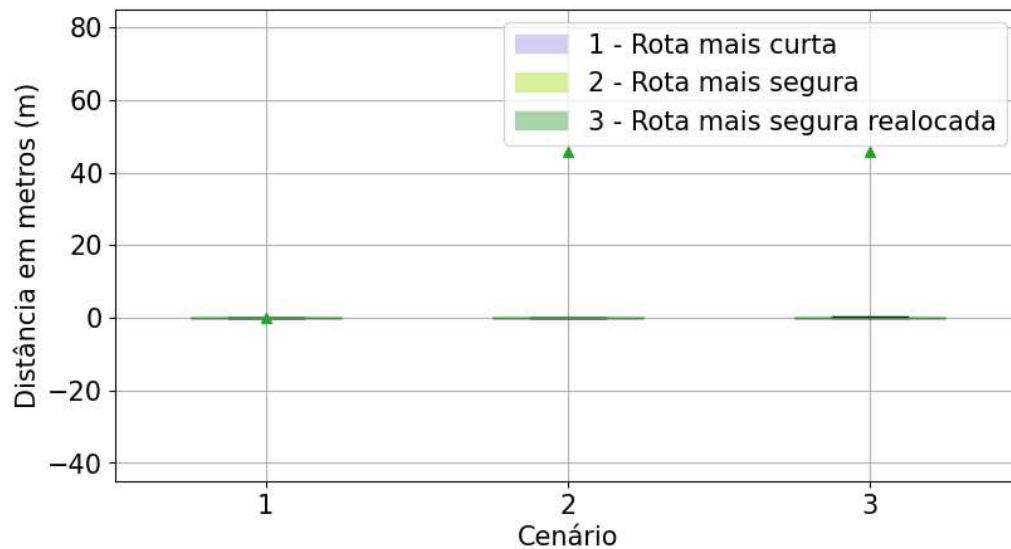


Figura 5.16: Acréscimo total do trecho na cidade de Curitiba.

Por fim, temos a **Distância dos nós em relação aos centros de clusters**. Percebe-se, que esta métrica também se manteve positiva para as duas cidades. Além disso, assim como no contexto de linhas completas, o ganho maior também se manteve no cenário de **ROTAS MAIS SEGURAS REALOCADAS**. Para São Paulo, o distanciamento médio foi de 5,34 metros para o cenário mais seguro, e 18,82 metros para o cenário mais seguro realocado, Figura 5.17.

Para Curitiba, a média do cenário mais seguro foi de 5,81 metros, e para cenário mais seguro realocado, o distanciamento médio foi de 13,43 metros, Figura 5.18. Percebe-se, que para as duas cidades as médias se mantiveram em intervalos próximos, entretanto, a diferença ainda sim ocorre, pois para São Paulo, a

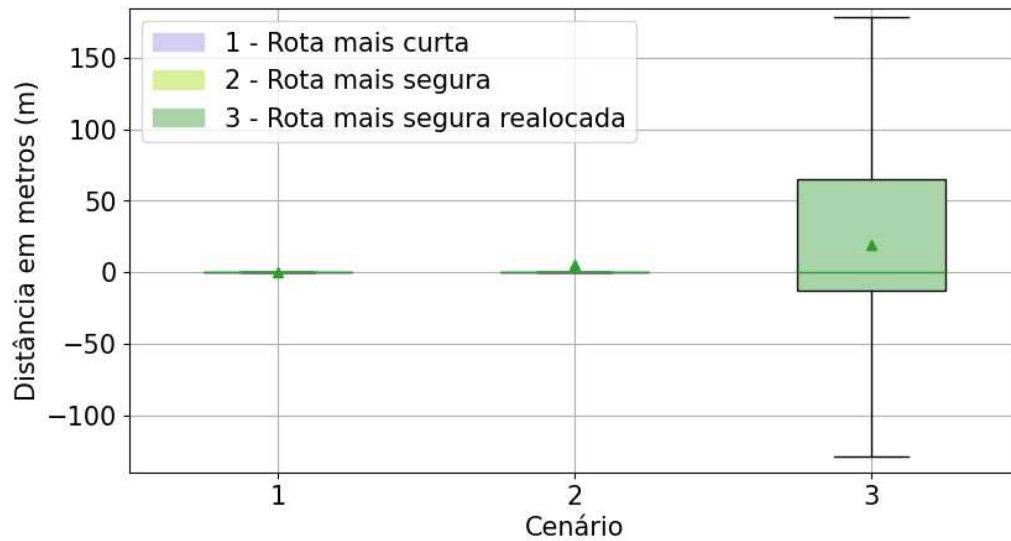


Figura 5.17: Distância dos nós em relação aos centros dos *clusters* para a cidade de São Paulo.

dificuldade para encontrar pontos de paradas disponíveis para realocação, e trajetos alternativos que se distanciem de fato de todos os *clusters* criminais próximos é muito maior, pois a cidade possui proporcionalmente muito mais regiões identificadas como inseguras, que acabam dificultando o trabalho do algoritmo em encontrar alternativas próximas do ponto de parada ou caminho original da rota. Por Curitiba não ser impactada dessa forma pelos polígonos, o algoritmo consegue identificar pontos de paradas e caminhos alternativos melhores, que culminam em trajetos sem desvios significativos, mas que ainda se distanciam significativamente dessas regiões. Por isso, é possível observar que para as métricas de comprimento de Curitiba, o acréscimo extra não possui diferença significativa entre os cenários com e sem realocação de pontos de paradas.

5.3 Avaliação da solução proposta com base em fatores socioeconômicos

Apesar dos critérios utilizados na construção da solução RACIONAL terem se mostrados satisfatórios nos resultados apresentados até o momento, ainda se mantêm os seguintes questionamentos: a distribuição dos polígonos criminais pelas cidades estão concentradas, em grande maioria, em regiões de classes sociais desfavorecidas? Além disso, existe algum tipo de viés em relação aos locais que receberam os maiores números de sugestões de troca?

Para responder a essas perguntas, foi elaborada uma nova análise, realizada em cima dos dados de rotas obtidos pela solução apresentada no trabalho. Por se tratar

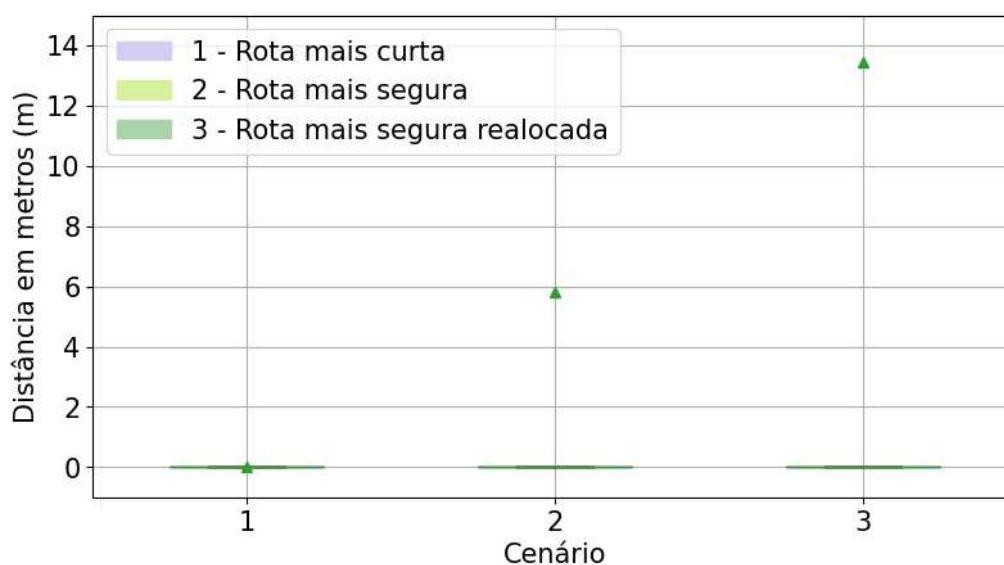


Figura 5.18: Distância dos nós em relação aos centros dos *clusters* para a cidade de Curitiba.

de uma abordagem que envolve dados socioeconômicos, foi necessário recorrer à base de dados pública do censo IBGE ³, que possui informações relacionadas à economia e distribuição de renda da população.

A partir deste dado bruto é possível aplicar filtros, transformá-los em estruturas de dados, e extrair as informações necessárias para o trabalho. Desta forma, a primeira etapa transformação envolve a representação dos setores censitários através de mapeamentos de polígonos da biblioteca Shapely ⁴ do Python. Esse dado, já representado dessa forma, contribuí para a integração com a solução construída no trabalho, tornando fácil a manipulação e agregação com a base de dados criada. Sendo assim, com essa informação, em conjunto da segunda, rótulo que representa a classe social econômica daquela região, já é possível conhecer a distribuição de renda contida nos *clusters* criminais.

5.3.1 Avaliando a distribuição socioeconômica dos polígonos criminais das cidades

Para obter este resultado, a primeira abordagem foi voltada à validação de todo o polígono estar contido no setor censitário, entretanto, a grande maioria dos polígonos não estavam totalmente contidos dentro das regiões dos setores, o que acabou gerando resultados imprecisos e construindo uma métrica incoerente. Para

³No Censo 2010, mais de 190 mil recenseadores visitaram 67,6 milhões de domicílios nos 5.565 municípios brasileiros. <https://censo2010.ibge.gov.br/resultados.html>

⁴Shapely: É um pacote Python licenciado pelo BSD para manipulação e análise de objetos geométricos planares. <https://shapely.readthedocs.io/en/stable/manual.html#polygons>

contornar este problema, foi necessário utilizar os centroides dos *clusters* criminais como referência, pois, por se tratar de apenas um ponto, este necessariamente estaria pelo menos dentro de um setor. Logo, ao final deste processo, já com todos os polígonos mapeados em seus respectivos setores censitários, foi preciso apenas agrupá-los por classe social para entender a relação estabelecida entre os dados socioeconômicos e as regiões identificadas como inseguras das cidades.

Nas análises realizadas, foram utilizados gráficos de Pareto para ilustrar a porcentagem de impacto de cada grupo no conjunto total de dados. Assim, a construção desses gráficos segue uma orientação que parte do grupo com o maior número de registros até o grupo com o menor, destacando o significativo impacto dos grupos maiores por meio da linha de porcentagem cumulativa, ao mesmo tempo em que também evidencia o baixo impacto dos grupos menores.

Começando por São Paulo, na Figura 5.19 é possível notar que aproximadamente 48% dos polígonos estão localizados como classe média, 30% como classe média alta, 20% como classe alta, e apenas 2% como classe baixa. Resultado que contempla cerca de metade (50%) dos polígonos localizados dentro de classes mais altas economicamente.

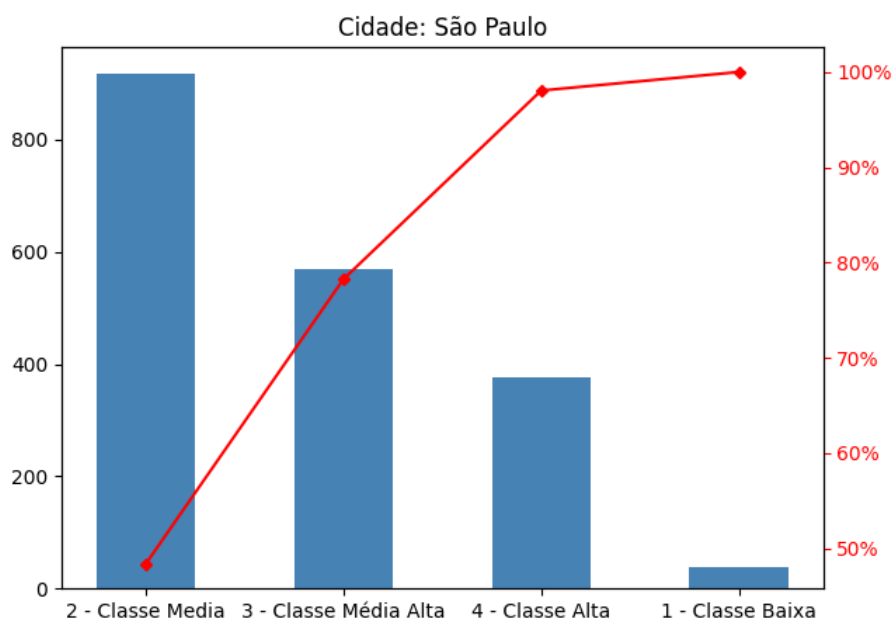


Figura 5.19: Gráfico de Pareto do agrupamento dos *clusters* criminais nos setores censitários para a cidade de São Paulo.

Já em Curitiba (Figura 5.20) percebe-se que 40% dos polígonos estão contidos na classe média, 37% como classe média alta, e 23% como classe alta, não possuindo dados para setores de classe baixa. Desta forma, assim como São Paulo, grande parte dos polígonos estão localizados dentro de regiões melhor situadas economicamente, neste caso, cerca de 60%.

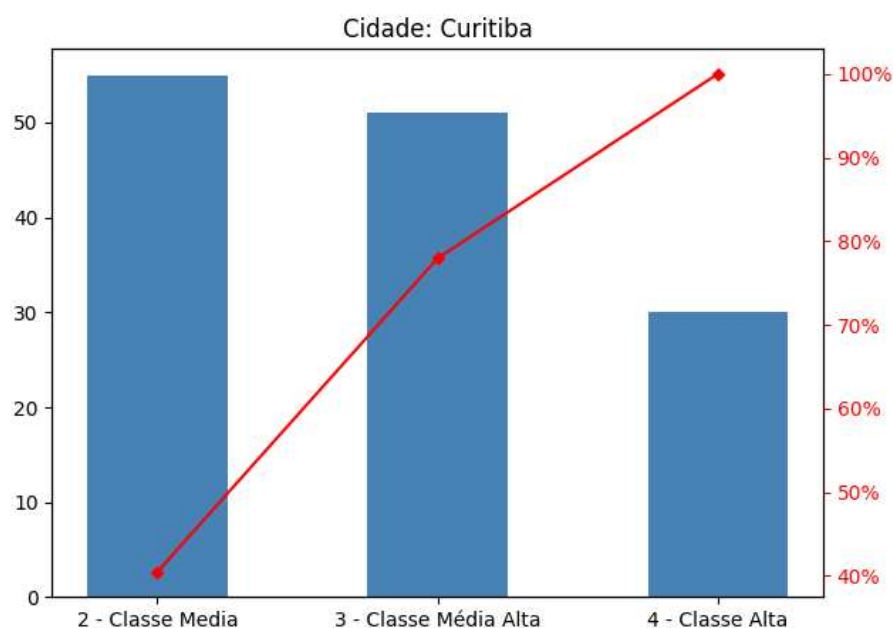


Figura 5.20: Gráfico de Pareto do agrupamento dos *clusters* criminais nos setores censitários para a cidade de Curitiba.

5.3.2 Avaliando os pontos de paradas e realocações feitas

Tendo mapeada a distribuição dos polígonos pelos setores censitários das duas cidades, uma outra métrica relevante para validar as perguntas levantadas no início dessa subseção estão relacionadas com as informações socioeconômicas dos pontos de paradas das duas cidades. Para criar essa análise de forma coerente com as modificações sugeridas pelo trabalho, dentre todos os pontos de paradas das duas cidades, apenas aqueles que receberam opções de realocação em pelo menos uma rota, e efetivamente foram substituídos, que serão considerados. Com isso, é possível mapear quais classes predominam estes pontos que foram substituídos, além também de ser possível fazer um levantamento das principais regiões das cidades que receberam tais modificações.

Para construir esta solução, o primeiro passo foi enriquecer os vértices do grafo que representam os pontos de parada realocados com as informações dos bairros e classes sociais. Durante este processo, foi possível mapear essas informações para praticamente toda a base, resultando em 2088 pontos de parada com esses dados completos (98,03%), enquanto apenas 42 (1,97%) apresentaram alguma informação ausente. Depois disso, foi feito um levantamento de todos os pontos de parada analisados (originais) e conseqüentemente substituídos, deixando o resultado já agrupado por número de ocorrências daquele mesmo ponto em diferentes rotas. Ao final deste processo, tem-se um *dataset* contando com todos os pontos de parada avaliados e outro com todos os substituídos, permitindo a comparação de ambos e a criação de métricas para cada etapa do processo.

Inicialmente são analisadas as modificações em relação aos pontos de paradas avaliados para realocação. Pois, é através dessa métrica que é possível ver a situação atual do sistema de transporte, localização destes dentro das classes sociais, e os principais bairros que receberam modificações. Na Figura 5.21 é possível perceber que grande parte dos pontos de paradas avaliados para troca estão dentro das classes médias, somando um total cerca de 78% dos pontos realocados. 20% representam a classe alta, restando então apenas 2% para a classe baixa.

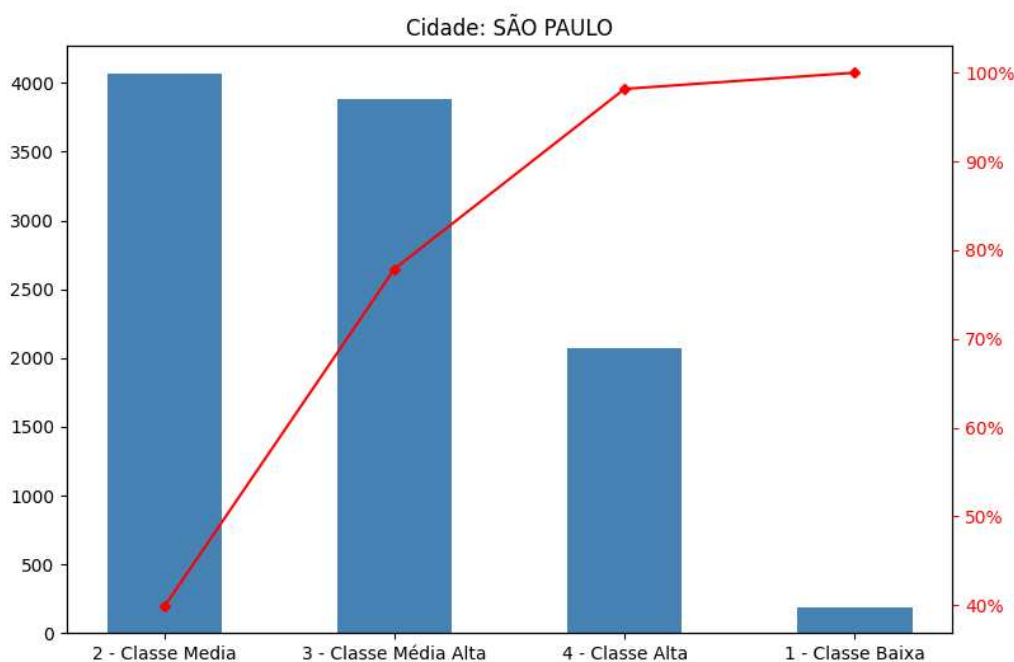


Figura 5.21: Gráfico de Pareto do agrupamento dos pontos de paradas avaliados que foram realocados na cidade de São Paulo.

Para Curitiba o resultado foi um pouco diferente. Na Figura 5.22 percebe-se que a grande parte dos pontos realocados estão dentre as classes melhor situadas economicamente, sendo aproximadamente 47% para classe alta, 31,5% para classe média alta, e 21,5% para a classe média.

Para se ter um panorama geral de como estão distribuídas e mapeadas todas as modificações feitas nas linhas originais, foram criadas as Tabelas 5.1 e 5.2. Através delas, é possível entender de forma mais abrangente o real impacto causado ao sugerir tais realocações nos sistemas de transportes públicos das duas cidades. Os resultados foram divididos de duas formas, para todas as realocações feitas, ou seja, as remoções dos pontos originais para os pontos candidatos, são agrupadas as alterações que mantiveram a mesma classe social, e as que foram trocadas.

Na cidade de São Paulo, Tabela 5.1, pode ser observado que quase 79% de todos os pontos realocados foram mantidos nas mesmas classes sociais que já pertenciam. Dentre as opções que sofreram mudanças, é possível perceber que existem diversas combinações de trocas, sendo possível tirar um ponto de uma região melhor

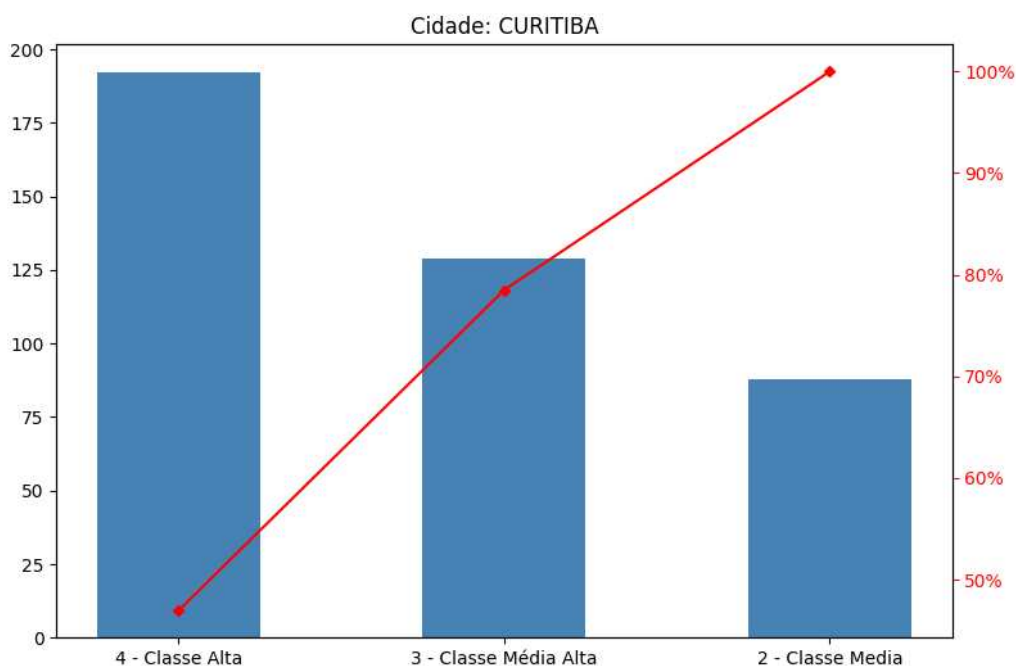


Figura 5.22: Gráfico de Pareto do agrupamento dos pontos de paradas avaliados que foram realocados na cidade de Curitiba.

localizada socialmente e levá-lo para uma de classe inferior, além do contrário também acontecer.

Classes Sociais		Quantidade	Total (Porcentagem)
Avaliada	Mantida		
2 - Classe Média	2 - Classe Média	593	1421 (78,64%)
3 - Classe Média Alta	3 - Classe Média Alta	444	
4 - Classe Alta	4 - Classe Alta	377	
1 - Classe Baixa	1 - Classe Baixa	7	
Avaliada	Substituída		386 (21,36%)
3 - Classe Média Alta	2 - Classe Média	116	
2 - Classe Média	3 - Classe Média Alta	100	
4 - Classe Alta	3 - Classe Média Alta	75	
3 - Classe Média Alta	4 - Classe Alta	66	
1 - Classe Baixa	2 - Classe Média	12	
2 - Classe Média	1 - Classe Baixa	7	
1 - Classe Baixa	3 - Classe Média Alta	5	
2 - Classe Média	4 - Classe Alta	4	
4 - Classe Alta	2 - Classe Média	1	

Tabela 5.1: Tabela com as métricas de impacto dos setores, causadas pelas realocações dos pontos de parada na cidade de São Paulo.

Para Curitiba, Tabela 5.2, cerca de 82% dos pontos realocados foram mantidos nas mesmas classes sociais, e dentre as modificações, os mesmos casos observados em São Paulo ocorreram nas mudanças de classes, reforçando a ideia de que o modelo

RACIONAL não possui relação direta com classes sociais quando propõe realocações de pontos de paradas. Além de também reforçar que o comprometimento estabelecido em não permitir alternativas fora do escopo da solução também funcionou de forma eficiente, mantendo ambas cidades com aproximadamente 80% das modificações mantidas nas mesmas classes sociais.

Classes Sociais		Quantidade	Total (Porcentagem)
Avaliada	Mantida		
4 - Classe Alta	4 - Classe Alta	109	229 (81,50%)
3 - Classe Média Alta	3 - Classe Média Alta	73	
2 - Classe Media	2 - Classe Media	47	
Avaliada	Substituída		
4 - Classe Alta	3 - Classe Média Alta	23	52 (18,50%)
2 - Classe Media	3 - Classe Média Alta	17	
3 - Classe Média Alta	2 - Classe Media	8	
3 - Classe Média Alta	4 - Classe Alta	4	

Tabela 5.2: Tabela com as métricas de impacto dos setores, causadas pelas realocações dos pontos de parada na cidade de Curitiba.

Além dos resultados apresentados que envolvem o impacto da solução RACIONAL proposta para as classes sociais, também foi feita uma análise em cima dessas realocações para o contexto dos bairros das cidades. Apesar da classe social representar de forma mais significativa e responder as perguntas que desencadearam os resultados já apresentados, o panorama criado pelos bairros nos permite entender de forma mais abrangente o real impacto da solução com as realocações propostas. Porém, apresentar este resultado se tornou um desafio, pois é uma métrica que envolve geolocalização e inúmeros bairros para ambas cidades, com isso, qualquer tipo de gráfico comum não representaria bem a geolocalização, além de tabelas também não condensarem bem os inúmeros bairros de forma coerente.

Sendo assim, para contornar essas dificuldades geoespaciais e de quantidade de dados, foi utilizada a abordagem de mapas de calor, adaptada para as duas cidades analisadas. Desta forma, é possível se situar melhor em cada bairro, além de comparar inúmeros dados através da variação da cor. O resultado desta métrica pode ser observado na Figura 5.23, para a cidade de São Paulo, e Figura 5.24 para Curitiba. São gráficos que contam com apenas dois conceitos principais, contidos nas legendas. O primeiro é o uso da tonalidade vermelha para representar as variações das quantidades de mudanças feitas em cada bairro, tendo os intervalos contidos nos extremos da legenda. Sendo assim, à medida que a cor escurece, mais modificações foram feitas naquele bairro. Entretanto, alguns bairros não receberam modificações. Com isso, para evitar interpretações confusas, foi utilizada a cor cinza para representá-los e separá-los dos demais que sofreram alterações.

Por fim, o resultado obtido é capaz de nos mostrar visualmente quais regiões e zonas das cidades que mais tiveram modificações. Em São Paulo é possível notar que os bairros com maiores números de realocações ficaram bem distribuídos, não ficando aglomerado em apenas um pedaço da cidade, além de ser possível notar o impacto causado pela remoção dos polígono gigante da região central da cidade. Com a manutenção dos mesmos, grande parte do centro estaria cinza, justamente por não possuir pontos de parada candidatos fora de *clusters* criminais. Já para Curitiba, as modificações ficaram mais concentradas na região oeste da cidade, mas ainda sim possuindo bairros com alterações significativas ao norte da cidade.

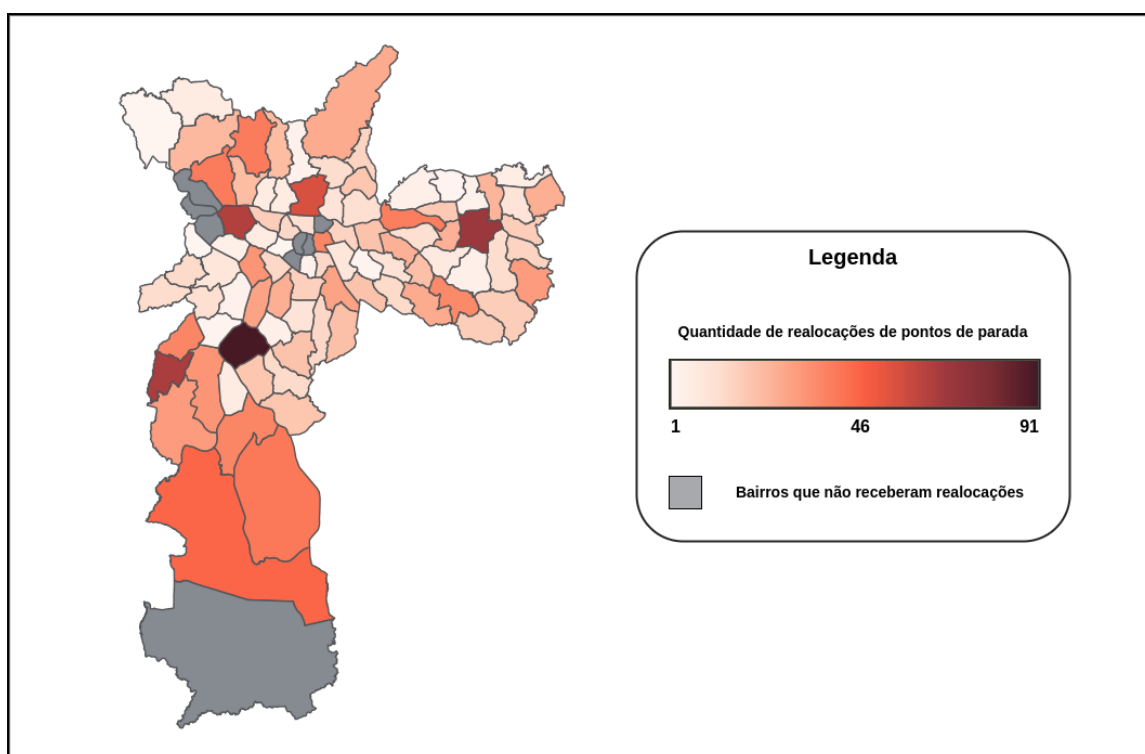


Figura 5.23: Mapa de calor dos bairros que mais receberam realocações na cidade de São Paulo.

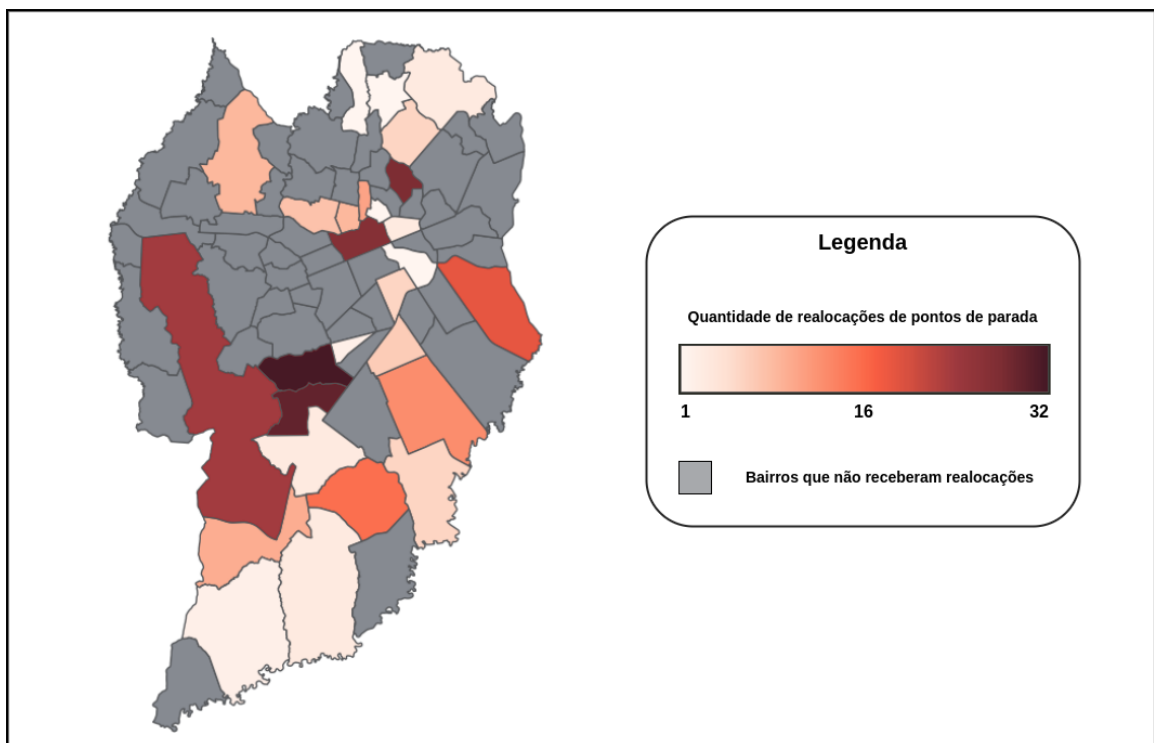


Figura 5.24: Mapa de calor dos bairros que mais receberam realocações na cidade de Curitiba.

Capítulo 6

Conclusão

Este trabalho apresentou a solução RACIONAL, modelo genérico que propõe uma proposta de elaboração de rotas seguras alternativas para ônibus urbanos como estratégia de mobilidade ciente de contexto. Como foi observado no decorrer do trabalho, se trata de uma área muito promissora e que tem muito potencial para resolver problemas de mobilidade urbana e questões sociais.

Como principais contribuições, neste trabalho foi possível criar uma solução genérica e centrada no usuário para a construção de rotas seguras, permitindo avaliar mais de uma cidade na implantação do modelo desenvolvido, e obter rotas seguras com ou sem mudança dos pontos de parada existentes.

Além disso, conseguimos desenvolver uma solução original, fundamentada em algoritmos da literatura, resultando em um modelo de fácil compreensão, customizável e aplicável a diversas cidades, independentemente de suas magnitudes. Tudo isso foi alcançado, possibilitando o equilíbrio entre rotas seguras, custos em relação à distância e o conforto dos usuários.

Por meio das métricas avaliadas, é possível notar como os cenários de rotas mais seguras apresentaram melhorias significativas em relação ao ganho de segurança se comparados aos cenários mais curtos. Resultados como a diminuição do deslocamento dentro de polígonos criminais, distanciamento dos pontos de paradas das regiões inseguras, e a quantidade de pontos de paradas que puderam ser realocados, refletem as inúmeras melhorias obtidas.

Além disso, o uso do cálculo de impedância e todos os outros critérios de realocação também serviram para validar a hipótese inicial da solução RACIONAL de construir rotas que priorizassem o conforto dos passageiros. O impacto disso pode ser observado nas métricas de impedância, pois, mesmo estabelecendo um limite fixo, os esforços médios ficaram aproximadamente 4x menores para São Paulo, e 10x menores para Curitiba.

Quanto aos pontos de parada modificados, observa-se que cerca de 80% desses pontos realocados, em ambas as cidades, permaneceram em suas respectivas classes sociais. Isso reforça a ideia de que o processo de realocação não apresenta viés relacionado a fatores socioeconômicos, até mesmo porque, dentre as situações onde

houveram as trocas de classes sociais, as realocações ocorreram tanto para classes sociais superiores quanto inferiores.

Todos esses ganhos apresentados puderam ser obtidos à um custo de aumentar o comprimento médio das rotas em aproximadamente 12%, no seu pior caso. Além disso, por meio dos dados de trânsito das cidades, é possível fazer um paralelo dessa distância extra com a realidade das cidades, resultando em um incremento de tempo de, no máximo, 11 minutos no pior caso. Isso demonstra que, apesar das alterações, essas não comprometeram significativamente o trajeto atual existente, ficando dentro dos limites estabelecidos, e ainda sim garantindo o acesso aos novos pontos de paradas sugeridos para os passageiros.

Como trabalhos futuros, uma possibilidade seria avaliar as magnitudes dos dados, e conseqüentemente dos resultados obtidos nas duas cidades, de modo a buscar alternativas para balanceá-los de forma mais equilibrada e equivalentes para ambas cidades. Isso pode ser interessante para avaliar combinações de parâmetros de clusterização mais próximos, gerando padrões de polígonos criminais com tamanhos mais balanceados e conhecidos. Além disso, também seria interessante avaliar a solução criada para cidades fora do Brasil, com diferentes padrões de malha rodoviária, relevo e mobilidade urbana.

Outra possibilidade seria aproveitar a flexibilidade do modelo para construir rotas seguras com base em novos dados e contextos, combinando aqueles já utilizados com elementos inéditos. Seguindo essa linha de raciocínio, seria interessante também adicionar critérios adicionais centrados nos usuários durante o processo de realocação de pontos de parada. Além disso, também relacionado à construção de rotas, uma possibilidade de otimização seria adaptar o método de construção de rotas para aproveitar o enriquecimento das arestas do grafo. Isso eliminaria a necessidade de gerar várias rotas e filtrá-las com base no índice de segurança K . Com essa abordagem, seria viável gerar as rotas mais curtas e seguras ajustando apenas o parâmetro de comparação do algoritmo, como no caso do algoritmo de Dijkstra, por exemplo.

Referências Bibliográficas

- Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., and Steggles, P. (1999). Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27–29, 1999 Proceedings 1*, pages 304–307. Springer.
- Al-Sultan, S., Al-Doori, M. M., Al-Bayatti, A. H., and Zedan, H. (2014). A comprehensive survey on vehicular ad hoc network. *Journal of network and computer applications*, 37:380–392.
- Almeida, V., Silva, T., and Silva, F. (2022). Se for, vá na paz: Construindo rotas seguras para veículos coletivos urbanos. In *Anais do XL Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 140–153, Porto Alegre, RS, Brasil. SBC.
- Aragão, J. (2006). Curso de gestão integrada da mobilidade urbana, módulo 01 - política nacional de mobilidade urbana. *Secretaria Nacional de Transporte e Mobilidade Urbana do Ministério das Cidades e Centro de Formação de Recursos Humanos em Transportes da Universidade de Brasília, Brasília, DF*.
- Babu, V. S. and Viswanath, P. (2008). An efficient and fast parzen-window density based clustering method for large data sets. In *International Conference on Emerging Trends in Engineering and Technology*, pages 531–536.
- Bachmann, C., Roorda, M. J., Abdulhai, B., and Moshiri, B. (2013). Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation. *Journal of Intelligent Transportation Systems*, 17(2):152–164.
- Bauza, R. and Gozávez, J. (2013). Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications. *Journal of Network and Computer Applications*, 36(5):1295–1307.
- Boarnet, M. G., Day, K., Anderson, C., McMillan, T., and Alfonzo, M. (2005). California's safe routes to school program: impacts on walking, bicycling, and pedestrian safety. *Journal of the American Planning Association*, 71(3):301–317.
- Bura, D., Singh, M., and Nandal, P. (2019). Predicting secure and safe route for women using google maps. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 103–108.

- Carvalho, C. H. R. d. (2016). Mobilidade urbana sustentável: conceitos, tendências e reflexões.
- Cavalcante, E. S., Cavalcante, L., Aquino, A., Pappa, G. L., and Loureiro, A. (2012). Uma abordagem evolutiva para posicionamento de pontos de disseminação em vanets. *XLIV Simpósio Brasileiro de Pesquisa Operacional (SBPO'12)*. IEEE.
- Crainic, T. G., Gendreau, M., and Potvin, J.-Y. (2009). Intelligent freight-transportation systems: Assessment and the contribution of operations research. *Transportation Research Part C: Emerging Technologies*, 17(6):541–557.
- Cunha, F., Maia, G., Celes, C., Guidoni, D., de Souza, F., Ramos, H., and Villas, L. (2017). Sistemas de transporte inteligentes: Conceitos, aplicações desafios. *Livro de Minicursos do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC'17)*, pages 59–103.
- de Souza, A. and Villas, L. (2020). Vem tranquilo: Rotas eficientes baseado na dinâmica urbana futura com deep learning e computação de borda. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 351–364, Porto Alegre, RS, Brasil. SBC.
- Denstadli, J. M. and Jacobsen, J. K. S. (2011). The long and winding roads: Perceived quality of scenic tourism routes. *Tourism Management*, 32(4):780–789.
- Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, 5:4–7.
- Dia, H. and Thomas, K. (2011). Development and evaluation of arterial incident detection models using fusion of simulated probe vehicle and loop detector data. *Information Fusion*, 12(1):20–27.
- Diab, E. I. and El-Geneidy, A. M. (2012). Understanding the impacts of a combination of service improvement strategies on bus running time and passenger's perception. *Transportation Research Part A: Policy and Practice*, 46(3):614–625.
- Félix Mata et al., 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.
- Gao, H. and Liu, F. (2013). Estimating freeway traffic measures from mobile phone location data. *European journal of operational research*, 229(1):252–260.

- Garg, T. and Kaur, G. (2022). A systematic review on intelligent transport systems. *Journal of Computational and Cognitive Engineering*, 2(3):175–188.
- HONGYANGA, W. (2013). Evaluation methodology of bus rapid transit (brt) operation. *Journal of Theoretical and Applied Information Technology*, 48(3):1733–1737.
- Kim, J. K., Sharman, R., Rao, H. R., and Upadhyaya, S. (2007). Efficiency of critical incident management systems: Instrument development and validation. *Decision Support Systems*, 44(1):235–250.
- 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.
- Lambert, J. H., Parlak, A. I., Zhou, Q., Miller, J. S., Fontaine, M. D., Guterbock, T. M., Clements, J. L., and Thekdi, S. A. (2013). Understanding and managing disaster evacuation on a transportation network. *Accident Analysis & Prevention*, 50:645–658.
- 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)*.
- Loke, S. (2006). *Context-aware pervasive systems: architectures for a new breed of applications*. CRC Press.
- Madkour, A., Aref, W. G., Rehman, F. U., Rahman, M. A., and Basalamah, S. (2017). A survey of shortest-path algorithms. *arXiv preprint arXiv:1705.02044*.
- Magagnin, R. C. and da Silva, A. N. R. (2008). A percepção do especialista sobre o tema mobilidade urbana. *Transportes*, 16(1).
- Naja, R. (2013). A survey of communications for intelligent transportation systems. In *Wireless vehicular networks for car collision avoidance*, pages 3–35. Springer.
- Natvig, M. K. and Vennesland, A. (2010). Flexible organisation of multimodal travel information services. *IET intelligent transport systems*, 4(4):401–412.

- Olaverri-Monreal, C., Gomes, P., Fernandes, R., Vieira, F., and Ferreira, M. (2010). The see-through system: A vanet-enabled assistant for overtaking maneuvers. In *2010 IEEE intelligent vehicles symposium*, pages 123–128. IEEE.
- Oliveira, E., Natarajan, M., and da Silva, B. (2019). Bus robberies in belo horizonte, brazil: Solutions for safe travel. *Crime & Delinquency*, 69(11):2359–2383.
- Oliveira, M. V. R. and Garcia, I. C. (2021). Sugestões de rotas personalizadas para carrinheiros na coleta seletiva de materiais recicláveis. In *Brazilian Symposium on Computer Networks and Distributed Systems*.
- Ozguven, E. E. and Ozbay, K. (2013). A secure and efficient inventory management system for disasters. *Transportation research part C: emerging technologies*, 29:171–196.
- Peng, Z.-R. (1997). A methodology for design of a gis-based automatic transit traveler information system. *Computers, environment and urban systems*, 21(5):359–372.
- Pires, C., Souza, W., Remboski, T., Aguiar, M., and Ferreira, P. (2014). Acoroute: Sistema de recomendação de rotas evitando congestionamento baseado em comunicação por feromônios. In *Anais do V Workshop de Computação Aplicada a Gestão do Meio Ambiente e Recursos Naturais*, pages 1–4, Porto Alegre, RS, Brasil. SBC.
- Poon, T., Choy, K. L., Chow, H. K., Lau, H. C., Chan, F. T., and Ho, K. (2009). A rfid case-based logistics resource management system for managing order-picking operations in warehouses. *Expert systems with applications*, 36(4):8277–8301.
- Qureshi, K. N. and Abdullah, A. H. (2013). A survey on intelligent transportation systems. *Middle-East Journal of Scientific Research*, 15(5):629–642.
- 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 sensoriados por usuários sobre crimes. In *Anais do XXII Workshop de Gerência e Operação de Redes e Serviços*, Porto Alegre, RS, Brasil. SBC.
- Shah, S. A. A., Kim, H., Baek, S., Chang, H., and Ahn, B. H. (2008). System architecture of a decision support system for freeway incident management in republic of korea. *Transportation Research Part A: Policy and Practice*, 42(5):799–810.
- 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.
- Vahdat-Nejad, H. (2014). Context-aware middleware: A review. *Context in Computing: A Cross-Disciplinary Approach for Modeling the Real World*, pages 83–96.

- Vahdat-Nejad, H., Ramazani, A., Mohammadi, T., and Mansoor, W. (2016). A survey on context-aware vehicular network applications. *Vehicular Communications*, 3:43–57.
- Wang, Y., Tan, G., Wang, Y., and Yin, Y. (2012). Perceptual control architecture for cyber-physical systems in traffic incident management. *Journal of Systems Architecture*, 58(10):398–411.
- Wedel, J. W., Schünemann, B., and Radusch, I. (2009). V2x-based traffic congestion recognition and avoidance. In *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks*, pages 637–641. IEEE.
- Winters, M., Babul, S., Becker, H., Brubacher, J. R., Chipman, M., Cripton, P., Cusimano, M. D., Friedman, S. M., Harris, M. A., Hunte, G., et al. (2012). Safe cycling: how do risk perceptions compare with observed risk? *Canadian journal of public health*, 103:S42–S47.
- Yan, Y., Guo, X., Li, Y., Kong, Z., and He, M. (2012). Bus transit travel time reliability evaluation based on automatic vehicle location data. *Journal of Southeast University*, 28(1):100–105.
- Yoon, S. W., Velásquez, J. D., Partridge, B., and Nof, S. Y. (2008). Transportation security decision support system for emergency response: A training prototype. *Decision Support Systems*, 46(1):139–148.
- Zhang, J., Liao, F., Arentze, T., and Timmermans, H. (2011). A multimodal transport network model for advanced traveler information systems. *Procedia-Social and Behavioral Sciences*, 20:313–322.
- Zumerchik, J., Rodrigue, J.-P., and Lanigan Sr, J. (2009). Automated transfer management systems and the intermodal performance of north american freight distribution. In *Journal of the Transportation Research Forum*, volume 48, pages 59–76.