

**ANÁLISE, INTEGRAÇÃO E VISUALIZAÇÃO DE  
DADOS URBANOS DE  
MÚLTIPLAS FONTES HETEROGÊNEAS**



PEDRO MAGALHÃES FORTINI

ANÁLISE, INTEGRAÇÃO E VISUALIZAÇÃO DE  
DADOS URBANOS DE  
MÚLTIPLAS FONTES HETEROGÊNEAS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais - Departamento de Ciência da Computação como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: CLODOVEU AUGUSTO DAVIS JÚNIOR

Belo Horizonte  
Dezembro de 2018

© 2018, Pedro Magalhães Fortini.  
Todos os direitos reservados.

Fortini, Pedro Magalhães

F742a Análise, integração e visualização de dados urbanos de múltiplas fontes heterogêneas / Pedro Magalhães Fortini. — Belo Horizonte, 2018  
xxiv, 96 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais - Departamento de Ciência da Computação

Orientador: Clodoveu Augusto Davis Júnior

1. Sistemas de Informação Geográfica.  
2. Computação Urbana. 3. Visualização de Dados.  
4. Análise Visual. I. Orientador. II. Título.

CDU 519.6\*04 (043)




UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

## FOLHA DE APROVAÇÃO

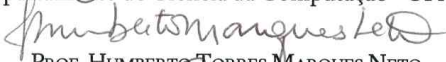
ANÁLISE, INTEGRAÇÃO E VISUALIZAÇÃO DE DADOS URBANOS  
DE MÚLTIPLAS FONTES HETEROGÊNEAS


**PEDRO MAGALHÃES FORTINI**

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

  
PROF. CLODOVEU AUGUSTO DAVIS JÚNIOR - Orientador  
Departamento de Ciência da Computação - UFMG

  
PROFA. MIRELLA MOURA MORO  
Departamento de Ciência da Computação - UFMG

  
PROF. HUMBERTO TORRES MARQUES NETO  
Departamento de Ciência da Computação - PUC/MG

  
PROF. RODRIGO AFFONSO DE ALBUQUERQUE NÓBREGA  
Departamento de Cartografia - UFMG

Belo Horizonte, 17 de dezembro de 2018.



*Dedico essa dissertação às pessoas mais presentes e importantes em minha vida:  
Aos meus pais e avós pelo amor incondicional.  
Aos meus irmãos por todo apoio e incentivo.  
Aos meus amigos pelas palavras e gestos de apoio nos momentos de dificuldade.*





# Agradecimentos

Agradeço primeiramente a Deus, por ter colocado pessoas tão especiais em minha vida, e por ser o meu suporte interior nos momentos mais difíceis.

Agradeço aos meus pais, Carlos Alberto e Guiomar, por me apoiarem em todas as minhas decisões, por todo investimento e incentivo em minha educação. Agradeço por sempre terem acreditado em minha capacidade e pelo amor incondicional.

Aos meus irmãos, Arthur e Ana, meu agradecimento especial, obrigado pela confiança, apoio e amizade.

Agradeço à minha avó Augusta pelo apoio e alegria trazidos nos momentos de dificuldade.

Faço um agradecimento especial ao meu orientador Prof. Clodoveu, por todo o conhecimento compartilhado, pelas idéias oferecidas, sem as quais seria impossível concluir este trabalho. Agradeço por todos os conselhos dados.

Agradeço aos meus amigos e familiares por toda palavra de incentivo, apoio e amizade. Enfim, gostaria de agradecer a todos que de alguma forma fizeram parte da minha vida no decorrer deste trabalho.



*“A persistência é o menor caminho do êxito”*  
(Charles Chaplin)



# Resumo

O rápido progresso do processo de urbanização e o conseqüente aumento da população urbana criam grandes desafios para os responsáveis pelo planejamento urbano. Esses desafios compreendem diversas áreas como meio-ambiente, consumo de energia, transporte, etc. A quantidade massiva de informações geradas todos os dias nos ambientes urbanos permite a realização de diagnósticos mais precisos de tais problemas e ajuda na criação de soluções. No entanto, como tais fontes de informação são heterogêneas e envolvem um grande volume de dados, existem grandes desafios computacionais para sua utilização. O presente trabalho visa estudar técnicas e métodos que permitam a análise, integração e visualização de dados urbanos de múltiplas fontes heterogêneas, objetivando criar um modelo genérico para análise de dados urbanos, com foco sobretudo em dados de transporte e trânsito. Para demonstrar a viabilidade, aplicabilidade e validade do modelo, foram criadas duas ferramentas: a *Impressão Digital de Trânsito Urbano* e o *Mapa Urbano de Acesso a Serviços Básicos*. A ideia principal da *Impressão Digital de Trânsito Urbano* é comparar o acesso e a mobilidade urbana em qualquer ponto de uma cidade utilizando as modalidades de transporte público (ônibus) e privado (individual). O *Mapa Urbano de Acesso a Serviços Básicos* foi criado com o objetivo de determinar a eficiência das rotas produzidas pelos sistemas de transporte público e privado de uma cidade, comparar a acessibilidade a serviços básicos utilizando tais sistemas de transporte, e contrastar tais informações com dados de qualidade de vida. Os resultados obtidos mostram grandes discrepâncias no esforço que um usuário de transporte público deve fazer para se locomover nas cidades analisadas em comparação com os modos individuais de transporte. Esses resultados ajudam a explicar as razões para a queda acentuada no número de usuários do transporte público nas cidades brasileiras nos últimos anos, conclamando os planejadores urbanos a criar políticas públicas na busca de melhores soluções.

**Palavras-chave:** Sistemas de Informação Geográfica, Computação Urbana, Visualização de Dados, Análise Visual.



# Abstract

The rapid progress of urbanization and the consequent increase in the urban population create great challenges for those responsible for urban planning. These challenges include various areas, such as the environment, energy consumption, transport, and others. Although there are many problems and challenges, the massive amount of information generated every day in urban environments allows envisioning solutions for them. However, as such sources of information are heterogeneous and involve a large amount of data, there are major challenges for their use. The objective of this dissertation is to study techniques and methods that allow the analysis, integration and visualization of urban data from multiple heterogeneous sources, aiming to create a generic model for urban data analysis, focusing mainly on transport and traffic data. In order to demonstrate the feasibility, applicability and validity of the model, two case studies were created: the *Urban Transit Fingerprint* and the *Urban Map of Access to Basic Services*. The main idea of the *Urban Transit Fingerprint* visualization is to compare the accessibility and urban mobility in any part of a city using public transportation (specifically buses) and private transportation. The *Urban Map of Access to Basic Services* was created with the objective of: determining the efficiency of the routes produced by the public and private transportation systems of a city, comparing accessibility to basic services using such transportation systems, and contrasting such information with data on quality of life. Results show wide discrepancies in the effort a public transit user has to make for her mobility in various regions of each city, as compared with individual modes of travel. These results help explaining the reasons behind the sharp drop in the number of transit users in Brazilian cities over the last years, thus calling on urban planners to devise public policies in the search for better solutions.

**Keywords:** Geographic Information Systems, Urban Computing, Visualization, Urban Visual Analytics.





# Lista de Figuras

2.1	Modelo geral para Computação Urbana . . . . .	6
2.3	Uso do <i>ThemeRiver</i> para representação linear do tempo . . . . .	11
2.2	Gráfico de linha apresentando a variação no número de viagens de táxi em 2011 e 2012 em Nova York . . . . .	11
2.4	<i>RingMap</i> mostrando variações de 96 atividades humanas (uma por anel) em um dia . . . . .	12
2.5	Precipitação nos Estados Unidos em um ano. Exemplo de uso da cor dos nós e a conexão entre eles para representação do tempo . . . . .	13
2.6	Relação entre o número de lojas de armas em cada distrito americano e o número de portadores de armas em cada estado . . . . .	14
2.7	<i>Heatmap</i> de um sistema de análise visual para a exploração de co-ocorrências na cidade de Guangzhou (China) . . . . .	15
2.8	Distribuição da população brasileira por estado no ano de 2012 . . . . .	15
2.9	Fluxos de migração nos Estados Unidos (Censo 2000), com mais de 3000 municípios e mais de 750.000 fluxos não-nulos de condado a condado . . . . .	16
2.10	200 principais fluxos de migração nos Estados Unidos suavizados . . . . .	16
2.11	Rede viária da cidade de Belo Horizonte . . . . .	17
2.12	Exemplo da técnica de agrupamento de arestas apresentada por Hurter et al. [16] . . . . .	17
2.13	Exemplo de visualização de dados textuais do tipo <i>Wordle</i> . . . . .	18
2.14	Cubo espaço-temporal apresentando a trajetória percorrida por uma pessoa em um dia na cidade de Enschede na Holanda . . . . .	18
2.15	Exemplo do uso de um sistema de coordenadas paralelas em um sistema de análise visual . . . . .	20
2.16	Visualização baseada em ícones desenvolvida por Cao et al. [5]. Utilizada para representar o comportamento de usuários em redes sociais . . . . .	21

2.17	Diagrama de nó-link criado por Zhao et al. [48] para representar a difusão de informações em mídias sociais . . . . .	21
2.18	<i>Nmap</i> que representa dez anos de artigos científicos da conferência IEEE Infovis e alguns outros artigos frequentemente citados em tal conferência . . . . .	22
2.19	Visualização presente na plataforma <i>LIVE Singapore!</i> . São apresentadas viagens e demandas de táxi em Cingapura durante fortes chuvas . . . . .	23
2.20	Visualização presente na plataforma <i>LIVE Singapore!</i> . Distorção da forma da cidade de Cingapura em um mapa para refletir o tempo que leva para viajar da origem selecionada para outras partes da ilha . . . . .	23
2.21	Menu inicial do <i>SocialGlass</i> , com várias opções de fontes e visualizações de dados . . . . .	24
2.22	Agrupamentos de pontos dinâmicos de micro-postagens do Twitter, em combinação com padrões de caminhamento de turistas estrangeiros e suas variações temporais na cidade de Roterdã . . . . .	24
2.23	Exemplo do diagrama <i>Urban Mobility Fingerprint</i> criado por Reimann et al. [34] para a cidade de Londres. . . . .	25
2.24	Visão geral do sistema criado por Liu et al. [24] para análise visual da diversidade de rotas . . . . .	26
2.25	Modelo de análise visual, aplicado a dados de transporte em larga escala proposto por Zhong et al. [53] . . . . .	26
2.26	Dois visualizações da ferramenta criada por Zhong et al. [53], a primeira apresenta a visão dos dados no formato de rede, e a segunda a visão geográfica dos dados . . . . .	27
2.27	Visualização de fluxos no nível de subzona na ferramenta criada por Zhong et al. [53] . . . . .	27
3.1	Modelo Geral para Computação Urbana. Foco em dados de transporte e trânsito . . . . .	30
3.2	Rede viária completa da região metropolitana da cidade de São Paulo e cidades próximas extraída utilizando o <i>OpenStreetMap</i> . . . . .	31
3.3	Polígono que define os limites do Município da cidade de São Paulo. . . . .	31
3.4	Rede Viária completa do Município da cidade de São Paulo . . . . .	33
3.5	Integração entre os dados da rede viária de uma cidade e dados de transporte público . . . . .	34
4.1	Rede viária completa da região metropolitana da cidade de Belo Horizonte e cidades próximas. . . . .	41

4.2	Polígono delimitador do Município da cidade de Belo Horizonte. . . . .	41
4.3	Rede Viária completa do município de Belo Horizonte . . . . .	41
4.4	Pontos dos <i>shapes</i> presentes no arquivo <i>shapes.txt</i> para a cidade de São Paulo. . . . .	47
4.5	Arestas da rede de itinerários de ônibus da cidade de São Paulo. . . . .	47
4.6	Pontos que constituem os <i>shapes</i> dos trajetos do sistema de transporte público da cidade de Belo Horizonte. . . . .	49
4.7	Rede de itinerários de ônibus da cidade de Belo Horizonte. . . . .	50
4.8	Rede de itinerários de ônibus da cidade de São Paulo. . . . .	50
4.9	Estrutura geral e relacionamentos das tabelas utilizadas para persistência dos dados das redes viária e de itinerários de ônibus . . . . .	51
4.10	Determinação do ângulo de rotação que alinhe uma rota verticalmente na direção de sul para norte . . . . .	53
4.11	Cálculo do fator de redimensionamento de uma rota na construção da <i>Impressão Digital de Trânsito Urbano</i> . . . . .	54
4.12	Resultados da aplicação da <i>Impressão Digital de Trânsito Urbano</i> para Belo Horizonte . . . . .	55
4.13	Divisão do município da cidade de Belo Horizonte nas suas regionais. . . . .	57
4.14	Pontos referentes aos serviços de saúde da cidade de Belo Horizonte. . . . .	58
4.15	Pontos referentes aos serviços de educação da cidade de Belo Horizonte. . . . .	58
4.16	Estrutura geral e relacionamentos das tabelas utilizadas para armazenamento dos dados das regionais, dos serviços de saúde e educação, e de qualidade de vida para a cidade de Belo Horizonte. . . . .	60
4.17	Projeção do ponto de origem posicionada entre o ponto de origem (A1) e destino (A2) da via. . . . .	62
4.18	Projeção do ponto de origem posicionada no ponto de origem (A1) da via. . . . .	62
4.19	Projeção do ponto de origem posicionada fora da via, antes do ponto de origem (A1) da via. . . . .	62
4.20	Projeção do ponto de destino posicionada entre o ponto de origem (A1) e destino (A2) da via. . . . .	63
4.21	Projeção do ponto de destino posicionada no ponto de destino (A2) da via. . . . .	63
4.22	Projeção do ponto de destino posicionada fora da via, após o ponto de destino (A2) da via. . . . .	63
4.23	Mapa coroplético da cidade de Belo Horizonte, as cores de cada regional indicam o valor do IDH da regional. . . . .	66
4.24	Resultados apresentados pelo sistema para a Regional Leste de Belo Horizonte. . . . .	66

5.1	Resultado final da <i>Impressão Digital de Trânsito Urbano</i> para o sistema de transporte privado tendo a Praça Sete de Setembro como origem das rotas	74
5.2	Resultado final da <i>Impressão Digital de Trânsito Urbano</i> para o sistema de transporte público tendo a Praça Sete de Setembro como origem das rotas	74
5.3	Visualizações produzidas para o sistema de transporte privado tendo a unidade da PUC-MG no bairro Coração Eucarístico como origem . . . . .	75
5.4	Visualizações produzidas para o sistema de transporte público tendo a unidade da PUC-MG no bairro Coração Eucarístico como origem . . . . .	75
5.5	Visualizações da <i>Impressão Digital de Trânsito Urbano</i> para o sistema de transporte privado tendo a Catedral Municipal de São Paulo como origem	76
5.6	Visualizações da <i>Impressão Digital de Trânsito Urbano</i> para o sistema de transporte público tendo a Catedral Municipal de São Paulo como origem .	76
5.7	Resultado final da <i>Impressão Digital de Trânsito Urbano</i> para o sistema de transporte privado com a Arena Corinthians como origem . . . . .	77
5.8	Resultado final da <i>Impressão Digital de Trânsito Urbano</i> para o sistema de transporte público com a Arena Corinthians como origem . . . . .	77

# Lista de Tabelas

5.1	Valores das medianas do <i>Indicador de Ineficiência da Rota</i> para os sistemas de transporte público e privado, em cada um dos pontos de interesse selecionados. . . . .	78
5.2	Tabela comparativa entre os dados de qualidade de vida (IDH) e o valor da mediana do <i>Indicador de Ineficiência do Transporte Público</i> por regional para os serviços de saúde de Belo Horizonte. . . . .	79
5.3	Tabela comparativa entre os dados de qualidade de vida (IDH) e o valor da mediana do <i>Indicador de Ineficiência do Transporte Público</i> por regional para os serviços de educação de Belo Horizonte. . . . .	80
5.4	Tabela comparativa entre o número de serviços na regional, as médias do tempo de viagem e da distância percorrida pelas rotas, e o valor da mediana do <i>Indicador de Ineficiência do Transporte Público</i> para os serviços de saúde de Belo Horizonte. . . . .	80
5.5	Tabela comparativa entre o número de serviços na regional, as médias do tempo de viagem ( $\bar{T}$ ) e da distância percorrida ( $\bar{D}$ ) pelas rotas, e o valor da mediana do <i>Indicador de Ineficiência do Transporte Público</i> para os serviços de educação de Belo Horizonte. . . . .	81



# Sumário

<b>Agradecimentos</b>	<b>ix</b>
<b>Resumo</b>	<b>xiii</b>
<b>Abstract</b>	<b>xv</b>
<b>Lista de Figuras</b>	<b>xvii</b>
<b>Lista de Tabelas</b>	<b>xxi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Objetivos . . . . .	3
1.3 Contribuições . . . . .	3
1.4 Publicações . . . . .	4
1.5 Estrutura do trabalho . . . . .	4
<b>2 Revisão Bibliográfica</b>	<b>5</b>
2.1 Computação Urbana: Visão Geral . . . . .	5
2.2 Análise visual em Computação Urbana . . . . .	8
2.2.1 Tipos de dados comumente utilizados . . . . .	8
2.2.2 Técnicas de visualização comumente utilizadas . . . . .	10
2.2.3 Visualização e abordagens de análise automatizada . . . . .	22
2.3 Considerações Finais . . . . .	27
<b>3 Metodologia</b>	<b>29</b>
3.1 Integração . . . . .	29
3.1.1 Extração dos dados brutos . . . . .	30
3.1.2 Pré-Processamento . . . . .	31
3.1.3 Integração e Padronização . . . . .	32

3.2	Armazenamento de dados . . . . .	34
3.3	Prestação de serviços . . . . .	35
3.3.1	Análise e Processamento . . . . .	35
3.3.2	Serviços . . . . .	36
3.4	Considerações Finais . . . . .	36
<b>4</b>	<b>Estudos de Caso</b>	<b>39</b>
4.1	Estudo de caso I: <i>Impressão Digital de Trânsito Urbano</i> . . . . .	39
4.1.1	Integração . . . . .	39
4.1.2	Armazenamento de dados . . . . .	50
4.1.3	Prestação de serviços . . . . .	51
4.2	Estudo de caso II: <i>Mapa Urbano de Acesso a Serviços Básicos</i> . . . . .	56
4.2.1	Integração . . . . .	56
4.2.2	Armazenamento de dados . . . . .	60
4.2.3	Prestação de serviços . . . . .	61
<b>5</b>	<b>Resultados</b>	<b>73</b>
5.1	Estudo de caso I: <i>Impressão Digital de Trânsito Urbano</i> . . . . .	73
5.2	Estudo de caso II: <i>Mapa Urbano de Acesso a Serviços Básicos</i> . . . . .	79
<b>6</b>	<b>Conclusão</b>	<b>85</b>
	<b>Referências Bibliográficas</b>	<b>89</b>
	<b>Anexo A Dados Utilizados</b>	<b>95</b>



# Capítulo 1

## Introdução

### 1.1 Motivação

O crescimento da população urbana decorrente do processo de urbanização afetou a vida de muitas pessoas, gerando no processo também grandes problemas e desafios para os responsáveis pelo planejamento dos grandes centros urbanos. Tais desafios abrangem diversos domínios, como transporte, consumo de energia e meio-ambiente, além de afetar as atividades diárias dos cidadãos. Associado a tal cenário temos muitos problemas, no entanto atualmente existem muitas fontes de informação que, a princípio, nos permitem construir soluções. Porém, como são fontes heterogêneas e envolvendo grandes volumes de dados, há grandes desafios para sua utilização.

Como indicado por Carvalho & Pereira [6], [7] e Gomide et al. [13], nas cidades brasileiras, a adoção de políticas e investimentos que priorizem o transporte privado sobre o transporte coletivo público provoca um aumento sistemático das tarifas. Com a redução do preço dos veículos particulares e o aumento da renda da população observado na última década, o aumento das tarifas de transporte público prejudica sua competitividade, resultando na perda de passageiros, que migram para modalidades de transporte individual, intensificando seu uso no dia a dia com forte impacto sobre as condições de mobilidade da população. Além disso, os serviços de compartilhamento de carros, como o Uber<sup>1</sup>, tornaram-se comuns, competindo com os modos tradicionais de transporte, como táxis e ônibus.

Os cidadãos decidem sobre alternativas de mobilidade com base em preferências pessoais, poder de compra e restrições de tempo. Portanto, qualquer tentativa de representar ou simular o funcionamento interno do sistema de trânsito deve levar em

---

<sup>1</sup><https://www.uber.com/pt-BR/br/>

conta todas as alternativas e suas características.

Segundo Zheng et al. [49], a *Computação Urbana* é um campo interdisciplinar no qual a Ciência da Computação se une a áreas de estudo como engenharia de transporte, engenharia civil, meio ambiente, economia, ecologia e sociologia visando a solução de problemas no contexto dos espaços urbanos. A *Computação Urbana* é uma área de estudo que aborda tais questões utilizando dados comumente gerados nas cidades, como fluxo de tráfego, ocorrências policiais e eventos de saúde pública. O principal objetivo de tal campo de estudo é ligar a detecção urbana, a gestão de dados, a análise de dados e a prestação de serviços num processo recorrente para uma melhoria contínua e não obstrutiva da vida das pessoas, dos sistemas de funcionamento da cidade e do meio ambiente.

Embora o termo Computação Urbana tenha sido cunhado e usado pela primeira vez em 2003 por Paulos & Goodman [31], e muitos pesquisadores como Castro et al. [8], Shklovski & Chang [37], Kostakos & O'Neill [20] e Zheng et al. [50] tenham trabalhado nisso ao longo dos anos, ainda existem algumas questões que não foram abordadas satisfatoriamente. Muitos trabalhos focam sobretudo no uso de metodologias e na construção de aplicações baseadas principalmente em abordagens automatizadas para mineração de dados. No entanto, como dito, a Computação Urbana é um campo de pesquisa multidisciplinar, portanto, o uso de uma análise completamente automática é difícil, e muitas vezes requer uma experiência considerável e profundo conhecimento em vários campos de pesquisa. Assim, é importante incluir a percepção humana no processo de exploração dos dados e combinar a flexibilidade, criatividade e conhecimento de domínio dos seres humanos com a enorme capacidade de armazenamento e poder computacional dos sistemas computacionais atuais.

Segundo [25], a visualização de dados é o estudo da transformação de dados e informações em representações visuais interativas e fornece uma maneira efetiva de integrar seres humanos em um processo de exploração de dados, aplicando suas habilidades perceptivas ao conjunto de dados alvo e alavancando seu conhecimento de domínio para orientar a exploração. Diante desse cenário, a análise visual visa combinar os métodos de análise automatizada com a visualização interativa de dados para uma compreensão, raciocínio e tomada de decisão efetivos com base em um conjunto de dados muito grande e complexo. Embora a importância da análise visual tenha sido reconhecida, especialmente para a Computação Urbana, esse continua sendo um conceito vago com muitas questões ainda pendentes como: analisar o papel da visualização de dados na Computação Urbana, avaliar como a visualização de dados pode ser combinada com algoritmos de mineração de dados e aprendizado de máquina, e identificar como funciona um sistema de análise visual urbana.

## 1.2 Objetivos

Diante do cenário apresentado, o presente trabalho tem como principal objetivo o estudo de técnicas e métodos que permitam a análise, integração e visualização de dados urbanos de múltiplas fontes heterogêneas permitindo a criação de um modelo genérico para análise de dados urbanos, com foco em dados de transporte e trânsito. Para tal, são analisadas técnicas para armazenagem e integração de dados de fontes heterogêneas, bem como algoritmos e visualizações que permitam realizar análises e extrair informações de tais dados.

Inicialmente, nos concentramos na formulação de um modelo genérico, que guia a construção de aplicações e sistemas em Computação Urbana, com foco em dados de transporte e trânsito. Após esse estágio inicial, e como forma de implementar e validar as diversas técnicas e metodologias estudadas e utilizadas na concepção do modelo, são apresentados estudos de caso, que na prática são aplicações para análise, integração e visualização interativa de dados urbanos, e que visam viabilizar e simplificar o uso de tais dados na tomada de decisões e no planejamento urbano. Os estudos de caso fornecem ferramentas para interpretação e compreensão da dinâmica urbana, utilizando o potencial combinado de fontes heterogêneas de dados de uma cidade.

## 1.3 Contribuições

As principais contribuições desse trabalho são:

- A criação do modelo genérico, que abarca os processos de obtenção, tratamento, integração e padronização dos dados, discute as melhores tecnologias de armazenamento de dados a serem utilizadas no cenário urbano, e apresenta a ideia de oferecer uma lógica separada entre os algoritmos e métricas criadas, e o serviço final oferecido para o usuário. Espera-se que tal modelo sirva como guia para outros trabalhos e projetos desenvolvidos na área.
- A criação de algoritmos para realização do *parsing*, tratamento e integração de dados do sistema de transporte público de uma cidade, fornecidos em um formato específico.
- A elaboração de um algoritmo que identifica a geometria das rotas que ligam estações de ônibus em sequência em um itinerário de ônibus, baseado na integração entre os dados dos itinerários de ônibus e da rede viária da cidade.

- A criação de novas métricas e índices que permitem a comparação das rotas produzidas por dois sistemas de transporte diferentes, nos casos estudados, os sistemas de transporte público e privado de uma cidade.
- A criação de novas ferramentas de visualização de dados, e a adaptação de técnicas de visualização já existentes para serem aplicadas no contexto de transporte e trânsito. A ideia é que tais ferramentas sejam utilizadas em conjunto com as métricas e índices criados, auxiliando no uso dos dados urbanos na tomada de decisões e no planejamento urbano.

## 1.4 Publicações

Como resultado desse trabalho foi publicado o artigo "*Analysis, Integration and Visualization of Urban Data From Multiple Heterogeneous Sources*" nos anais do *1st International Workshop on Advances in Resilient and Intelligent Cities (ARIC)* parte da conferência *26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2018)*. No artigo são apresentados o modelo genérico criado, a aplicação de tal modelo na construção da ferramenta de visualização discutida na Seção 4.1, e os resultados do uso dessa ferramenta com os dados de transporte das cidades de Belo Horizonte e São Paulo, que indicam a validade do uso da ferramenta no processo de tomada de decisão sobre o planejamento urbano.

## 1.5 Estrutura do trabalho

Esta dissertação está organizada em seis capítulos. O segundo capítulo possui uma revisão bibliográfica, no qual são tratados os principais trabalhos e conceitos relacionados a Computação Urbana, visualização de dados urbanos, e análise visual em Computação Urbana. No terceiro capítulo apresentamos o modelo genérico para análise de dados urbanos, criado com foco em dados de transporte e trânsito. O quarto capítulo apresenta dois estudos de caso nos quais o modelo é utilizado na criação de serviços que auxiliam a tomada de decisões e o planejamento urbano. No quinto capítulo mostramos resultados obtidos com o uso dos serviços criados, e como tais resultados podem de fato auxiliar os planejadores urbanos. O sexto capítulo apresenta a conclusão do trabalho, avaliando o mesmo, e apresentando possíveis linhas de pesquisa que podem ser desenvolvidas a partir dessa dissertação.

# Capítulo 2

## Revisão Bibliográfica

Neste capítulo é apresentada uma revisão bibliográfica acerca dos temas abordados neste trabalho. Foi realizada uma pesquisa de temas pertinentes à Computação Urbana bem como os principais projetos relacionados a cada um deles. O entendimento dos conceitos e temas abordados é necessário para a compreensão e desenvolvimento das análises presentes nesse trabalho.

### 2.1 Computação Urbana: Visão Geral

Apesar de ser um campo de pesquisa muito recente, existe um conjunto de projetos em Computação Urbana com propósitos relevantes e que se relacionam com o trabalho proposto.

Zheng et al. [49] construíram um modelo geral para Computação Urbana, tal modelo é apresentado na Figura 2.1. O modelo é composto de quatro camadas: detecção urbana, gerenciamento de dados urbanos, análise de dados e fornecimento de serviços. Usando a detecção de anomalias urbanas feita por Pan et al. [30] como exemplo, o modelo funciona da seguinte forma:

- Na etapa de detecção urbana, a mobilidade das pessoas é constantemente medida (por exemplo, as rotas utilizadas na rede rodoviária de uma cidade) usando sensores de GNSS (*Global Navigation Satellite Systems*) ou seus sinais de telefone celular. Também são coletados continuamente dados de redes sociais.
- Na etapa de gerenciamento de dados, a mobilidade e os dados das redes sociais são organizados utilizando alguma estrutura de indexação que simultaneamente incorpora informações espaço-temporais e dados textuais para permitir uma análise de dados eficiente.

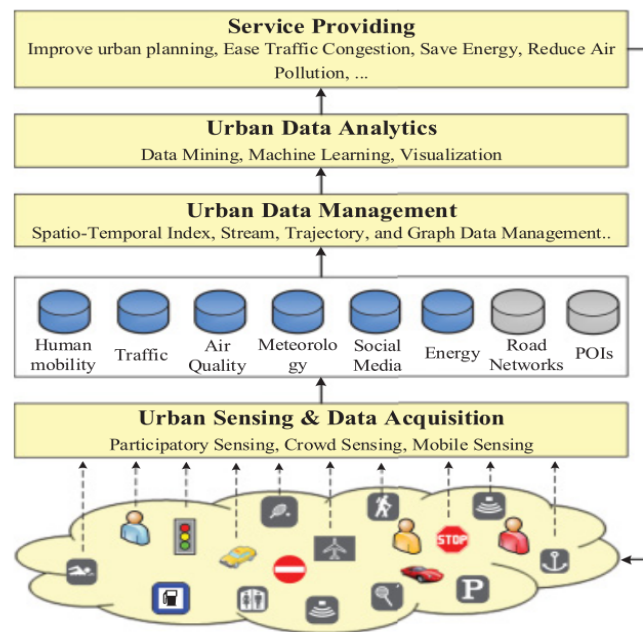


Figura 2.1: Modelo geral para Computação Urbana. Fonte: Zheng et al. [49].

- Na etapa de análise de dados, uma vez que ocorra uma anomalia, é possível identificar os locais onde a mobilidade das pessoas difere significativamente de seus padrões de origem. Enquanto isso, é possível também descrever a anomalia através da mineração de termos representativos das mídias sociais que estão relacionados aos locais e horário.
- No passo do fornecimento de serviços, os locais e a descrição da anomalia são enviados para motoristas próximos, para que possam escolher uma derivação no caminho que estão seguindo. Além disso, a informação é entregue a autoridades de transporte, para que possam dispersar o tráfego e diagnosticar a anomalia. O sistema continua a execução, permitindo uma detecção instantânea e discreta de anomalias urbanas, ajudando a melhorar as experiências de condução das pessoas e reduzir os congestionamentos na cidade.

Comparado com outros sistemas (por exemplo, motores de busca na Web) que se baseiam em um modelo de tarefa única (ou seja, recuperação de informações de páginas da Web), a Computação Urbana possui uma estrutura multi-modal, ou multitarefa. Dessa forma, diferentes tarefas podem ser cumpridas combinando diferentes fontes de dados com diferentes técnicas de aquisição, gerenciamento e análise de dados de diferentes camadas do modelo.

Zheng et al. [49] apresentam as principais aplicações em Computação Urbana, mostrando diversas possibilidades e trabalhos que utilizaram os conceitos estudados na

área. Mais especificamente, são definidas sete categorias de aplicações: Planejamento Urbano, Sistema de Transporte, Meio-ambiente, Consumo de Energia, Aplicações e Mídia Social, Economia e Segurança Pública. Focamos sobretudo nas categorias de Planejamento Urbano, Sistema de Transporte e Aplicações e Mídias Sociais, que estão mais relacionadas com o presente trabalho.

A formulação do planejamento urbano requer a avaliação de uma ampla gama de fatores, como fluxo de tráfego, mobilidade, pontos de interesse e estruturas de transporte. Esses fatores complexos e de rápida evolução transformam o planejamento urbano em uma tarefa muito desafiadora. Liu et al. [27] tentam solucionar os problemas subjacentes na rede de transportes de Pequim, analisando as trajetórias de GNSS geradas por 33.000 táxis durante um período de 3 anos. Já Yuan et al. [46] propõem um framework para definir regiões de diferentes funções em uma cidade, utilizando dados de mobilidade entre regiões e pontos de interesse localizados nas mesmas. Outro problema para o planejamento urbano é a definição dos limites entre as diversas regiões de uma cidade. A definição dessas fronteiras de acordo com a interação entre as pessoas pode fornecer ferramentas de apoio para a tomada de decisões. Ratti et al. [33] propõem uma abordagem refinada para delimitação regional através da análise de uma rede deduzida a partir de uma grande base de dados de telecomunicações na Grã-Bretanha. Dada uma área geográfica e alguma medida da força das ligações entre os seus habitantes, eles dividem essa área em regiões menores e não sobrepostas. O algoritmo gerou regiões geograficamente coesas que correspondem às regiões administrativas.

Sistemas que encontram as melhores rotas são importantes tanto para reduzir o tempo de deslocamento dos motoristas, quanto para reduzir o consumo de combustível. Thiagarajan et al. [39] apresentam o VTrack, um sistema para estimativa do tempo de viagem baseado em sinais WiFi, medindo e localizando possíveis atrasos. O sistema usa um esquema de correspondência de mapas utilizando cadeias de Markov e baseado em dados intermitentes para identificar os segmentos de vias mais prováveis utilizados pelo usuário. Propõe-se então um método de estimação do tempo de viagens para atribuir tempos de deslocamento aos segmentos. Além da sugestão de rotas, a Computação Urbana tem papel importante na melhoria do sistema de transporte urbano. Nesse sentido, Zimmerman et al. [55] criaram o Tiramisu, um sistema no qual os passageiros compartilham trilhas de GNSS, recolhidas a partir dos receptores de seus telefones. O Tiramisu então processa os dados de entrada e gera previsões de tempo de chegada em tempo real dos ônibus. Também no campo da melhoria do transporte público, Lathia et al. [22] utilizaram dados coletados do sistema de metrô de Londres para construir planejadores de rotas de viagem mais precisos. Para tal, foram realizadas análises sobre dois grandes conjuntos de dados do uso do metrô de Londres, que demonstraram

que existem diferenças substanciais entre os viajantes. Com base nessas idéias, eles extraíram características dos dados que implicitamente capturam informações sobre a familiaridade de um usuário com uma viagem, a similaridade de um usuário com outros viajantes e o contexto de viagem de um usuário. Finalmente, eles usaram tais recursos para desenvolver ferramentas personalizadas de planejamento de viagens.

Na análise de redes sociais, os trabalhos em geral focam nas redes sociais baseadas em localização. Tais redes funcionam como ponte entre o comportamento dos usuários nos mundos digital e físico, que corresponde bem à natureza da Computação Urbana. Um exemplo de aplicação nesse sentido é apresentado por Zheng et al. [52], que criaram um grafo bipartido entre usuários e locais e empregaram um modelo baseado em HITS (*Hypertext-Induced Topic Search*, Pesquisa de Tópicos Induzidos por Hipertexto) para inferir o nível de interesse de um local e o conhecimento de viagem de um usuário. A ideia geral é que as experiências de viagem dos usuários e o nível de interesse de um local têm uma relação de reforço mútuo.

## 2.2 Análise visual em Computação Urbana

Zheng et al. [51] fazem uma revisão das principais técnicas, conceitos e projetos de análise visual no contexto da Computação Urbana. Os autores apresentam dados mais utilizados em Computação Urbana, as técnicas de visualização mais utilizadas para cada tipo de dado, e como a visualização de dados pode ser combinada com abordagens de análise automatizadas.

### 2.2.1 Tipos de dados comumente utilizados

Zheng et al. [51] categorizam os tipos de dados urbanos usados com frequência na Computação Urbana em cinco categorias:

1. **Dados de mobilidade urbana:** É o tipo de dado mais utilizado em análises urbanas, pode auxiliar no estudo das dinâmicas social e comunitária. Com base em diferentes fontes de dados, podem ser categorizados em dados de tráfego, dados de deslocamento, dados de telefonia móvel e dados de redes sociais georreferenciadas.
  - **Dados de tráfego:** Tipo de dados gerados e coletados por sensores em veículos (por exemplo, táxis, ônibus, metrô, trens, embarcações e aviões) ou monitores instalados ao longo das estradas (por exemplo, sensores e câmeras de vigilância).



- **Dados de deslocamento:** Dados que registram o movimento regular das pessoas nas cidades. Entre os vários tipos de dados estudados, os de cartões de transporte público são um exemplo típico desse tipo. Hoje em dia, em uma cidade moderna, os passageiros costumam usar cartões personalizados nos leitores de cartões em ônibus ou estações de metrô para entrar e sair do sistema de transporte público, gerando assim uma grande quantidade de registros de viagens de passageiros. Esse tipo de dado pode ser utilizado não só para melhoria do sistema de transporte público de uma cidade, como também para analisar padrões de mobilidade urbana.
  - **Dados de telefonia móvel:** São os registros de todas as conexões (por exemplo, chamadas telefônicas, mensagens, internet) entre telefones celulares e estações de celular coletadas por operadores de telecomunicações. Além das informações de comunicação, esse tipo de dado fornece a localização dos usuários com base na localização das estações de celular, o que oferece recursos importantes para o estudo da mobilidade urbana.
2. **Dados de mídias sociais:** As mídias sociais se tornaram uma forma de comunicação extremamente popular, gerando assim, uma grande quantidade de dados. Esses dados, contém informação valiosa, permitindo estudar as relações entre diferentes pessoas, bem como a estrutura de uma determinada comunidade. Além disso, dados gerados por mídias sociais como textos, fotos e vídeos contém informações preciosas sobre os interesses e características de um usuário, o que pode auxiliar em pesquisas como a evolução da opinião pública sobre determinados temas nas mídias sociais, e a divulgação de informações anômalas.
  3. **Dados geográficos:** Tipo de dados fundamental na análise visual urbana que fornece estrutura básica, bem como informações semânticas para cenários de Computação Urbana. São dados usados com frequência desse tipo:
    - **Rede Viária:** Normalmente representada por um grafo no qual as arestas representam os segmentos de vias e os nós representam as interseções de tais segmentos. Os nós são representados como pontos geograficamente localizados, enquanto as arestas podem ter informações como comprimento, limite de velocidade e número de pistas.
    - **Rede de Transporte:** São as rotas de trânsito e pontos de parada das redes de ônibus e metrô. Cada ponto de parada possui um identificador e suas coordenadas geográficas. Além disso, informações sobre os itinerários também são incluídas, como o quadro de horários que mostra quando cada

ônibus ou metrô sai do terminal de partida e chega a cada parada ao longo de sua rota.

- **Pontos de interesse:** São dados relacionados a instalações como restaurantes, shopping centers, parques, aeroportos, escolas e hospitais da cidade. Cada instalação geralmente é descrita por um nome, endereço, categoria e um conjunto de coordenadas geográficas.

4. **Dados ambientais:** A aceleração do processo de urbanização fez surgir sérios problemas ambientais em todo o mundo, como poluição ambiental e consumo de energia cada vez maior. Muitos estudos foram dedicados a analisar conjuntos de dados relacionados a tais problemas. Tais conjuntos de dados podem ser categorizados em duas classes: dados de monitoramento do meio ambiente e dados de consumo de energia. O primeiro inclui dados meteorológicos (por exemplo, temperatura, umidade, duração do sol e condições climáticas) como os apresentados por Li et al. [23], dados de poluição do ar, como os usados por Chung et al. [9], dados da qualidade da água, como aqueles usados por Accorsi et al. [1] e dados de sensoriamento por satélite, usados por Wu et al. [44]. O último registra dados como consumo de eletricidade, como é feito por Goodwin et al. [14], e gás, presente no projeto de Mühlbacher & Piringer [29], o que pode ajudar a avaliar e otimizar o uso de energia em uma cidade, detectando correlações e prevendo possíveis picos de demanda.
5. **Outros dados relacionados:** As cidades modernas são unidades integradas e abrangentes, logo outras fontes de dados não necessariamente conectadas diretamente ao contexto urbano também podem ser utilizadas nas análises em Computação Urbana. Klemm et al. [18] e Borkin et al. [4] utilizam dados de saúde, Zhang et al. [47] utilizam dados de serviço público, Xie et al. [45] e Ziegler et al. [54] desenvolvem projetos com dados econômicos, Trimm et al. [41] e Shi et al. [36] utilizam bases de dados de educação, e Dietrich et al. [10] e Janetzko et al. [17] fizeram trabalhos com dados de eventos esportivos.

Com o desenvolvimento da computação e das técnicas de gerenciamento e análise de dados, mais fontes de dados urbanos estarão disponíveis, o que nos permitirá modelar uma cidade melhor e mais inteligente no futuro próximo.

### 2.2.2 Técnicas de visualização comumente utilizadas

No contexto da Computação Urbana as três propriedades fundamentais extraídas de dados urbanos são tempo (quando), espaço (onde) e objeto (o que). Esses são compo-

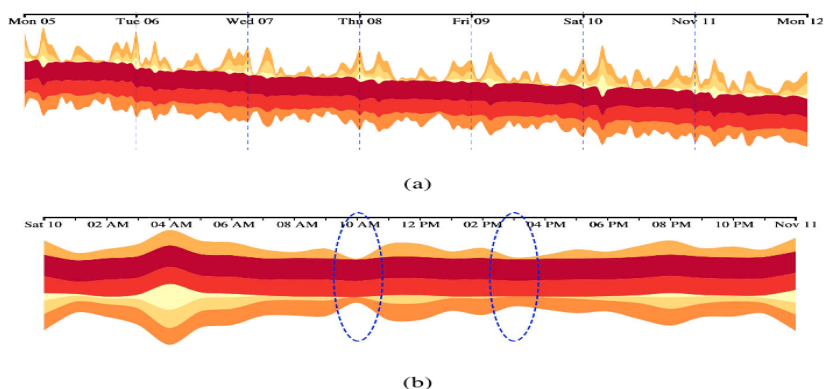


Figura 2.3: Uso do *ThemeRiver*, uma forma de gráfico empilhado, para representar o tempo linear [44]. A figura é uma visualização dos dados de acessibilidade do centro da cidade de Xangai: (a) Dados do período de 5 de novembro a 11 de novembro de 2007 mostra diferentes padrões temporais para diferentes dias (separados por linhas pontilhadas); (b) Período de 10 de novembro de 2007 (sábado) com dois vales às 10 da manhã e às 15h. marcado com azul indica dois picos de tráfego.

nentes elementares para a apresentação de fatos ocorridos no contexto urbano e podem ajudar a estruturar o domínio da informação.

### 2.2.2.1 Visualização do tempo

O tempo nesse caso é considerado como a medida pela qual os dados urbanos podem ser ordenados do passado através do presente para o futuro. Consideramos a medida das durações de tempo e intervalos entre itens de dados como propriedades numéricas. Existem várias maneiras de mapear o tempo para variáveis visuais. No contexto urbano, a mais comum é o design baseado em eixos. Para visualização do tempo linear, a visualização mais utilizada são gráficos, como o gráfico de linhas (Figura 2.2) e o gráfico empilhado (Figura 2.3).

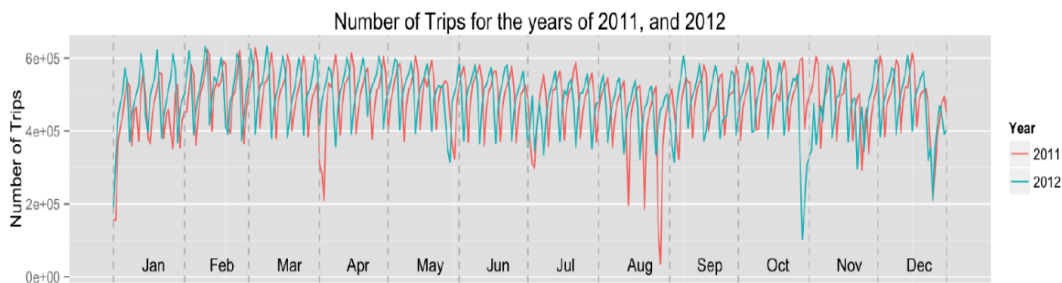


Figura 2.2: Gráfico de linha que representa o tempo linear [12]. A figura mostra como o número de viagens de táxi varia em relação a 2011 e 2012 em Nova York

Em alguns casos, a ideia é representar atividades que acontecem de maneira

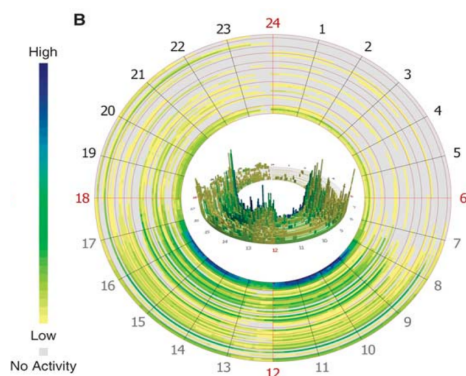


Figura 2.4: *RingMap* mostrando variações de 96 atividades humanas (uma por anel) em um dia [2]. É apresentada uma representação 3D que facilita na identificação de picos nos dados

periódica. Nesses casos, um eixo circular similar ao de um relógio é utilizado, como apresentado na Figura 2.4. A cor e a conexão são dois métodos comuns para interpretar o tempo em dados urbanos. Temos exemplo desse uso nos trabalhos de Wu et al. [44] e Bach et al. [3]. Exemplos de apresentação de ordem cronológica de dados usando cor e conexão são apresentados na Figura 2.5. Estes dois métodos, embora sejam amplamente utilizados, têm uma limitação relacionada à sua escalabilidade, causada pela capacidade limitada dos olhos humanos de distinguir diferentes cores e a potencial confusão de conexões densamente agrupadas. Portanto, eles são melhor utilizados em combinação com interações que permitam a visualização de detalhes dos dados. Além de visualizações estáticas, a informação temporal também pode ser transmitida através de uma representação visual dinâmica, o que resulta em visualizações que mudam ao longo do tempo automaticamente (por exemplo, animações). No entanto, como Robertson et al. [35] demonstram, animações geralmente não são formas de visualização de dados eficazes para tarefas analíticas, devido ao fato de que a memória humana de curto prazo é limitada.

### 2.2.2.2 Visualização do espaço

A maior parte dos dados urbanos são coletados com suas localizações geográficas. Tais localizações no contexto urbano são geralmente especificadas de três formas: georreferenciamento baseado em coordenadas, georreferenciamento baseado em divisão e georreferenciamento linear. Cada uma dessas especificações possui um tipo de visualização associado, tais tipos são apresentados a seguir:

- **Visualização baseada em pontos:** É o tipo de visualização comumente utilizada para dados com georreferenciamento baseado em coordenadas de objetos

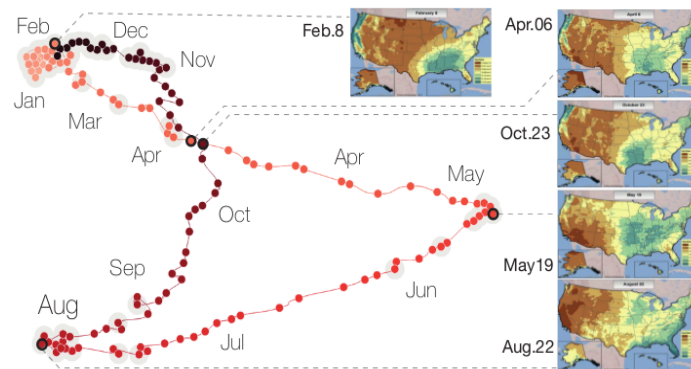


Figura 2.5: Precipitação nos Estados Unidos em um ano [3]. A visualização utiliza a cor dos nós e a conexão entre eles para indicar o tempo

estáticos ou posições instantâneas de objetos em movimento. A ideia básica é colocar pontos individualmente no contexto espacial (por exemplo, em um mapa). Cada ponto representa um objeto ou evento, e os canais visuais (por exemplo, cor e tamanho) desses pontos codificam informações relacionadas (por exemplo, status de objetos ou categoria de eventos). Como exemplo, a Figura 2.6 apresenta um mapa dos Estados Unidos demonstrando a relação entre o número de lojas de armas em cada distrito e o número de portadores de armas em cada estado. Nele o tamanho dos pontos representando os distritos indica o número de lojas de armas no mesmo. Visualizações baseadas em pontos permitem aos usuários observar claramente objetos ou eventos de forma individual nos dados. No entanto, quando o número de objetos ou eventos se torna grande, a grande concentração de pontos torna a visualização pouco clara e difícil de interpretar. Os *Heatmaps*, como o apresentado na Figura 2.7 são uma solução comum para enfrentar esse problema.

- **Visualização baseada em regiões:** A visualização baseada na região é frequentemente usada para mostrar informações agregadas com base em regiões de uma divisão predeterminada do espaço. O mapa Coroplético apresentado na Figura 2.8 é um típico exemplo de visualização baseada em região, que mostra regiões como polígonos e codificam um atributo (no caso, a população de cada estado brasileiro no ano de 2012) utilizando variações de cores. Além disso, para visualizar os fluxos entre regiões, o mapa de fluxo pode ser utilizado como apresentado por Speckmann et al. [38]. No entanto, um mapa de fluxo facilmente se torna ilegível devido às interações maciças e sobreposição de fluxos à medida em que o tamanho dos dados aumenta, como pode ser visto na Figura 2.9. Guo & Zhu [15] propuseram uma nova abordagem para o alinhamento e mapeamento de

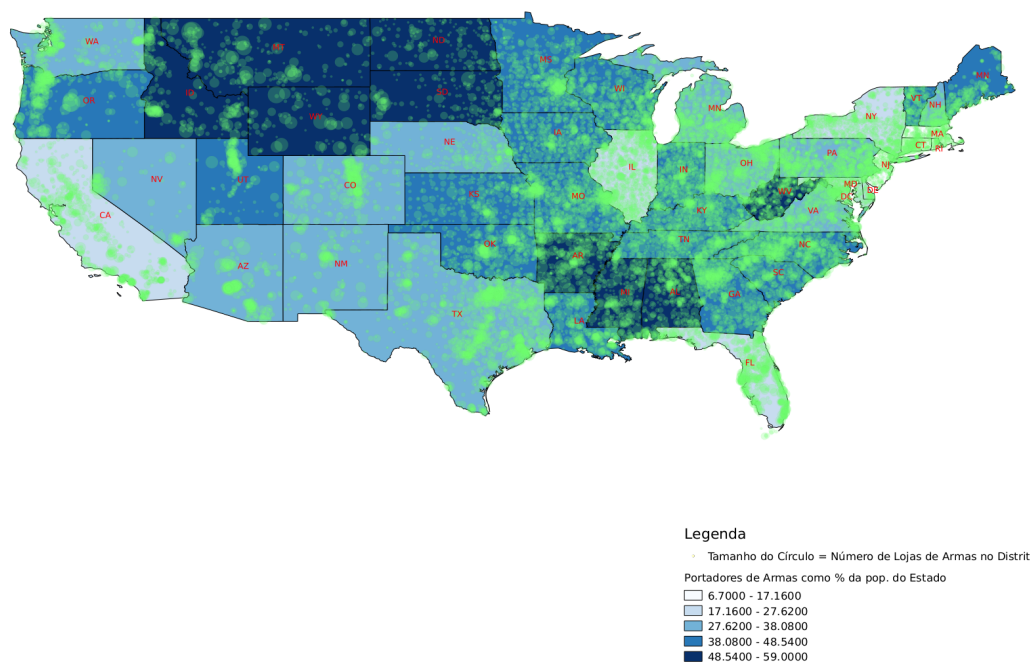


Figura 2.6: Relação entre o número de lojas de armas em cada distrito americano e o número de portadores de armas em cada estado. Como indicado, cada ponto no mapa é um distrito nos Estados Unidos, e o tamanho dos pontos representa o número de lojas de armas no distrito

fluxo com base em uma nova definição de fluxo vizinho. Um exemplo de mapa de fluxo suavizado para padrões de migração nos EUA é mostrado na Figura 2.10. Nela vemos os mesmos dados apresentados na Figura 2.9, porém aplicando a técnica de suavização de fluxo desenvolvida por Guo & Zhu [15]. Em geral, a visualização baseada em regiões tem vantagens para revelar padrões macro (por exemplo, fluxos entre regiões), entretanto é inadequada para analisar padrões micro (por exemplo, o comportamento de um indivíduo). Portanto, esse tipo de técnica é frequentemente usado em combinação com outras técnicas para suportar uma análise abrangente com diferentes níveis de detalhe.

- Visualização baseada em linhas:** No contexto urbano, localizações podem ser especificadas com base em mapas rodoviários ou redes viárias. A visualização baseada em linha pode ser usada para apresentar esse tipo de localização. Um exemplo é dado na Figura 2.11, na qual é apresentada a rede viária da cidade de Belo Horizonte. Com a melhoria das tecnologias de posicionamento, é possível transformar pontos de dados discretos em estruturas contínuas chamadas trajetórias, que podem permitir a identificação de mais informações semânticas no espaço. A visualização baseada em linha também pode ser usada para retratar

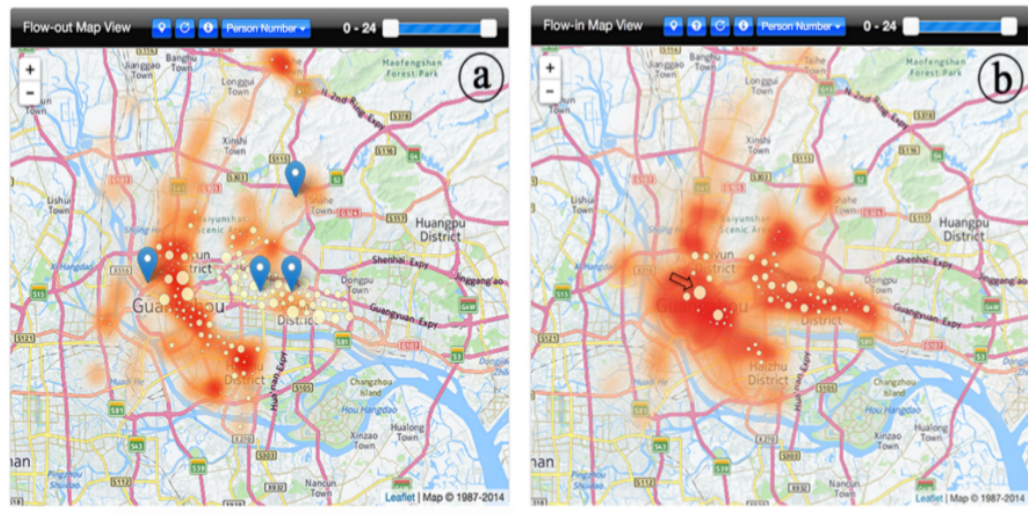


Figura 2.7: *Heatmap* de um sistema de análise visual [43] para a exploração de co-ocorrências (ou seja, pessoas de duas regiões diferentes visitam determinado local urbano durante o mesmo período) no contexto de mobilidade urbana na cidade de Guangzhou (China), o sistema utiliza dados de telecomunicações para realizar suas análises

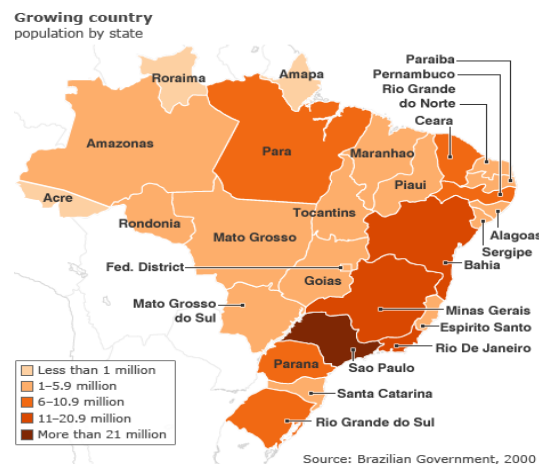


Figura 2.8: Distribuição da população brasileira por estado no ano de 2012

localizações com base em trajetórias. Convencionalmente, uma trajetória pode ser desenhada como uma linha ou curva em um mapa que se conecta do ponto inicial ao seu ponto final. Muitas abordagens foram desenvolvidas para evitar a desordem visual que ocorre quando temos um grande número de trajetórias. O agrupamento de arestas é uma das abordagens mais populares para realizar tal tarefa. Um exemplo de aplicação do agrupamento de arestas na visualização baseada em linha é mostrado na Figura 2.12, que apresenta a técnica de agrupamento usada por Hurter et al. [16]. Embora o agrupamento de arestas seja

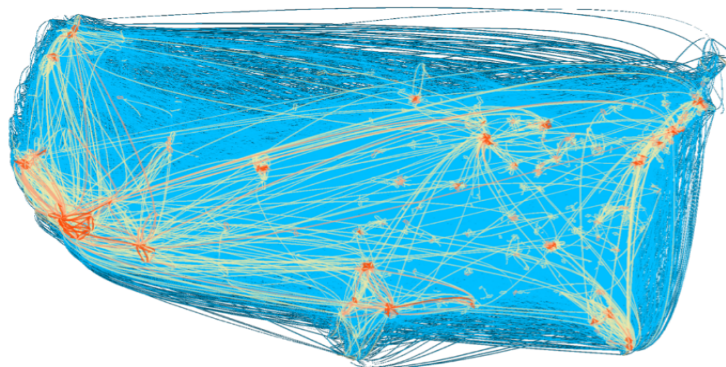


Figura 2.9: Fluxos de migração nos Estados Unidos (Censo 2000), com mais de 3000 municípios e mais de 750.000 fluxos não-nulos de condado a condado. Este mapa mostra fluxos com 100 ou mais migrantes, o que representa 13% da migração nos dados

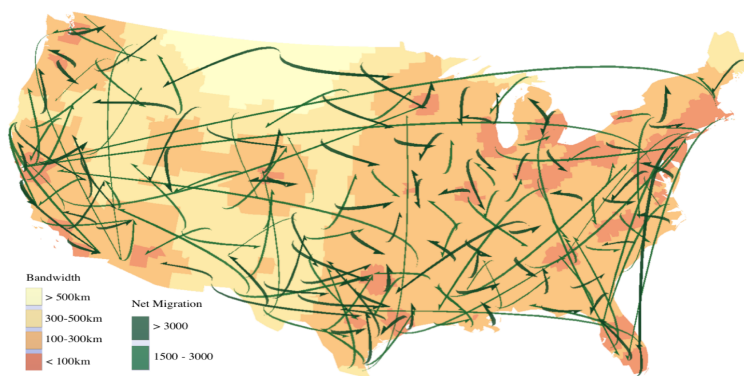


Figura 2.10: 200 principais fluxos de migração nos Estados Unidos suavizados [15]. O tamanho da população considerada é  $p = 1.000.000$ . Tanto a largura da linha quanto a cor representam valores de fluxo. O mapa de fundo mostra a largura do fluxo de cada município

efetivo na redução da desordem visual, ele introduz ambiguidades visuais que podem impedir a compreensão de trajetórias. Logo, deve ser utilizado em conjunto com formas interativas de apresentar detalhes nos dados.

### 2.2.2.3 Visualização de outras propriedades

Os dados urbanos possuem outras informações além das espaciais e temporais anteriormente discutidas. Essas propriedades podem ser categorizadas em propriedades numéricas, propriedades categóricas e propriedades textuais. As propriedades numéricas referem-se a medidas de magnitude que permitem a comparação aritmética, enquanto as propriedades categóricas são atributos discretos que podem auxiliar a determinar se dois objetos de dados são iguais ou diferentes. Munzner [28] apresenta estudos que exploram canais visuais adequados para codificar propriedades numéricas e categóricas.



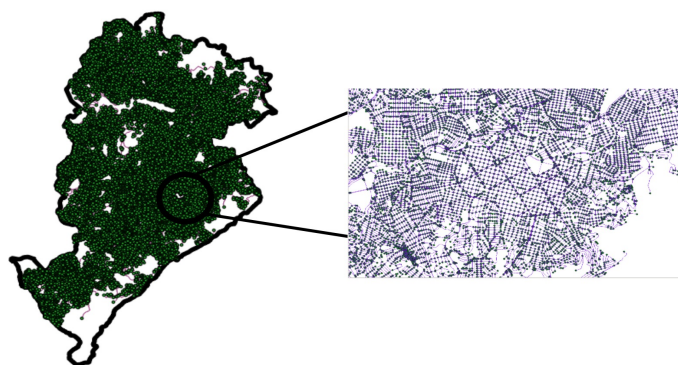


Figura 2.11: Rede viária da cidade de Belo Horizonte. As linhas representam segmentos de vias, e os nós interseções entre esses segmentos

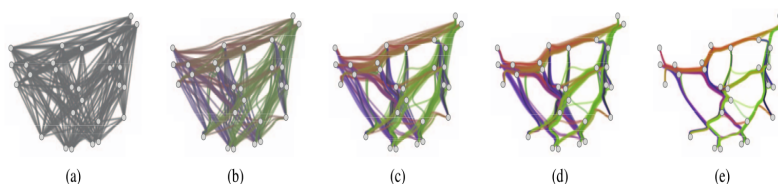


Figura 2.12: Exemplo da técnica de agrupamento de arestas apresentada por Hurter et al. [16]. A cor é utilizada para indicar a direção das arestas

Segundo Munzner [28], o mesmo atributo de dados codificado com dois canais visuais diferentes resultará em interpretação diferente pelo sistema perceptivo humano. Portanto, é necessário orientar o design para a visualização de dados com base na classificação efetiva de canais visuais para propriedades numéricas e categóricas.

Propriedades textuais referem-se a informações gravadas em formato de texto, como conteúdo de postagens na rede social e nome de segmentos rodoviários. Essas propriedades fornecem informações semânticas ricas, essenciais para análises e interpretações aprofundadas. Técnicas de visualização de texto, como a nuvem de termos do *Wordle*, apresentado na Figura 2.13, podem ser empregadas para a visualização de propriedades textuais. Da mesma forma, técnicas de processamento de linguagem natural também podem ser integradas para a extração de informações e transformar propriedades textuais em outros tipos de dados.

#### 2.2.2.4 Visualização de múltiplas propriedades

Na análise de dados urbanos para aplicações do mundo real, muitas vezes é necessário considerar simultaneamente as **propriedades espaciais e temporais dos objetos**. O cubo espaço-temporal é uma técnica típica para a visualização simultânea de espaço e tempo. A ideia dessa técnica é mapear as duas dimensões espaciais através de dois eixos

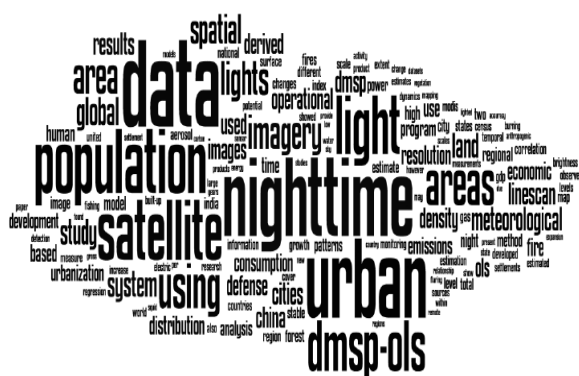


Figura 2.13: Exemplo de visualização de dados textuais do tipo *Wordle*

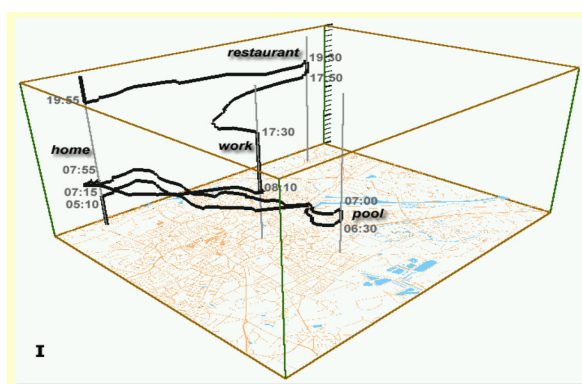


Figura 2.14: Cubo espaço-temporal apresentando a trajetória percorrida por uma pessoa em um dia na cidade de Enschede na Holanda

(isto é, os eixos  $x$  e  $y$ ) de um cubo virtual de três dimensões e usar a terceira dimensão (isto é, o eixo  $z$ ) para o mapeamento do tempo. O contexto espacial é muitas vezes representado por um mapa que constitui uma face do cubo. Assim, pode-se colocar objetos gráficos no cubo para marcar eventos de interesse, ou construir trajetórias que ilustram o caminho dos objetos. Um exemplo de cubo espaço-temporal é apresentado na Figura 2.14. Em muitos casos, é necessário alterar o cubo espaço-tempo padrão para descrever atributos associados, como no trabalho desenvolvido por Tominski et al. [40]. No entanto, o uso da visualização 3D é controverso e pode levar a ambiguidades devido a problemas perceptivos como potenciais oclusões de dados. Portanto, para uma análise efetiva, o cubo espaço-tempo geralmente necessita ser utilizado em conjunto com formas de interação apropriadas para permitir que os usuários vejam os dados de diferentes perspectivas.

Muitas outras propriedades também estão envolvidas na análise de dados urbanos. Na maior parte do tempo, tais propriedades estão correlacionadas e a análise dessas correlações é essencial para aplicações do mundo real. Portanto, a visualização de múltiplos

tiplas propriedades simultaneamente é necessária. Em comparação com a visualização espaço-temporal, as técnicas de **visualização de propriedades multivariadas** podem ser empregadas para visualizar dados urbanos de forma mais compacta, o que é particularmente importante quando é necessária a comparação de um número grande de propriedades. Liu et al. [26] argumentam que as técnicas de visualização de propriedades multivariadas podem ser classificadas em quatro categorias: técnicas baseadas em *pixels*, projeções geométricas, técnicas baseadas em ícones e técnicas hierárquicas. A seguir discutimos e apresentamos exemplos de visualizações que utilizam tais técnicas:

- **Técnicas baseadas em *pixels*:** A idéia básica é mapear valores de dados para *pixels* com um esquema de cores adequado e, em seguida, agrupar os *pixels* adequadamente para serem utilizados em tarefas analíticas. A forma mais comum de organizar os *pixels* é utilizando matrizes de *pixels*, devido à sua interpretabilidade e simplicidade, como é feito por Wang et al. [42] e Krueger et al. [21]. Além disso, o segmento do círculo é uma variante da forma da matriz, cuja ideia é representar os dados em um círculo dividido em segmentos, um para cada atributo. O *Ringmap* apresentado na Figura 2.4 é um exemplo dessa técnica.
- **Projeções geométricas:** O objetivo é encontrar projeções informativas e transformações para dados multivariados, de modo a usar a posição espacial dos objetos para codificar tais dados visualmente. Devido à complexidade dos dados urbanos, encontrar uma projeção apropriada não é uma tarefa fácil. Uma possibilidade é usar diagramas de dispersão, nos quais os itens de dados são mapeados ao longo dos eixos x e y de um sistema de coordenadas cartesianas definido por duas propriedades de dados ou projeções multidimensionais, como a análise de componentes principais (PCA) ou dimensionamento multidimensional (MDS). A segunda opção é usar um gráfico de coordenadas paralelas que representa um item de dados multivariados como uma polilinha com vértices em eixos paralelos e cada eixo corresponde a uma propriedade de dados. Um exemplo é apresentado na Figura 2.15. No entanto, as coordenadas paralelas sofrem de duas limitações principais: como determinar a ordem dos eixos, uma vez que diferentes formas de ordenação destacam diferentes aspectos da estrutura de dados, e o problema da desordem visual causado por um grande número de polilinhas.
- **Técnicas baseadas em ícones:** Visa fornecer uma forma compacta de mapear as propriedades de cada item de dados multivariados para recursos gráficos, como um ícone ou glifo. Para a análise visual urbana, existem muitos exemplos de uso de ícones, especialmente em conjunto com outras visualizações, como mapas,

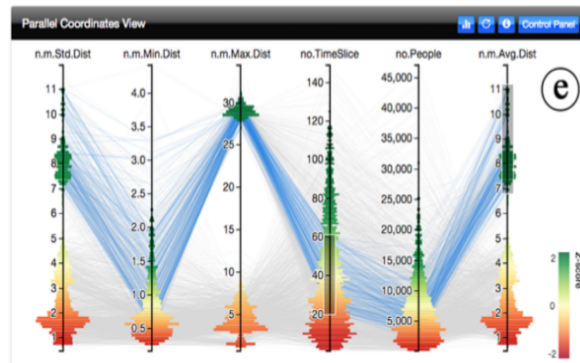


Figura 2.15: Exemplo do uso de um sistema de coordenadas paralelas em um sistema de análise visual. A visualização está presente no sistema desenvolvido por Wu et al. [43] para investigação de co-ocorrências. A visualização de coordenadas paralelas permite uma análise quantitativa eficiente baseada em uma ampla gama de atributos

diagramas de dispersão, linhas de tempo e gráficos, o que facilita a execução de uma análise abrangente. É amplamente aceito que um bom design de um ícone pode facilitar a aprendizagem, a memorização e a comparação eficazes, enquanto um design de um ícone pouco efetivo pode provocar confusão perceptiva, ambiguidade semântica ou dificuldade de aprender e lembrar. Em geral, não é uma tarefa fácil projetar um ícone efetivo, que seja fácil de entender e memorizar. Embora a maioria dos projetos existentes tenha sofrido um processo duradouro de evolução, refinamento e padronização, uma das críticas mais comuns ao uso de ícones é que existe um viés implícito na maioria dos mapeamentos (ou seja, algumas características gráficas ou relações entre recursos são mais difíceis de perceber do que outras). Um exemplo de visualização por ícone, desenvolvido por Cao et al. [5], é apresentada na Figura 2.16. A ideia é auxiliar na identificação de usuários com comportamentos anômalos em sistemas de comunicação online, como emails e mídias sociais.

- **Técnicas hierárquicas:** Técnicas projetadas para visualizar a estrutura hierárquica dos dados, incluindo hierarquias de dimensão e hierarquias de dados. Essa técnica divide o espaço de dados em subespaços e organiza-os de forma hierárquica. Para visualização de dados urbanos, **Diagramas de nó-link** e **Treemaps** são duas formas visuais usadas com frequência. Em um diagrama de nó-link, os itens de dados são apresentados como nós com ligações entre nós que representem relacionamentos correspondentes. Existem várias aplicações de diagramas de nó-link no contexto urbano, incluindo visualização de comunidades, análise de tópicos e a interpretação da difusão de informações, a Figura 2.17 apresenta

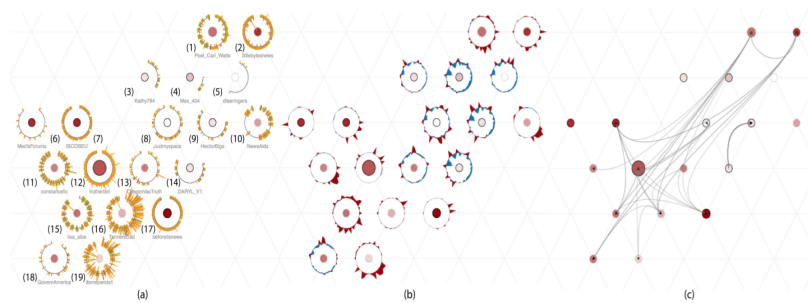


Figura 2.16: Visualização por ícone desenvolvida por Cao et al. [5]. Nesta visualização, os usuários rotulados de 1 a 19, são representados como círculos dimensionados por sua importância (ou seja, seu número de seguidores), coloridos por suas pontuações de anomalia variando de branco (menor) a vermelho escuro (maior). Em diferentes modos de visualização, cada círculo é cercado por uma visualização de (a) os segmentos de atividade do usuário, ou (b)  $z$ -score das características do usuário, ou (c) links que indicam as interações entre usuários

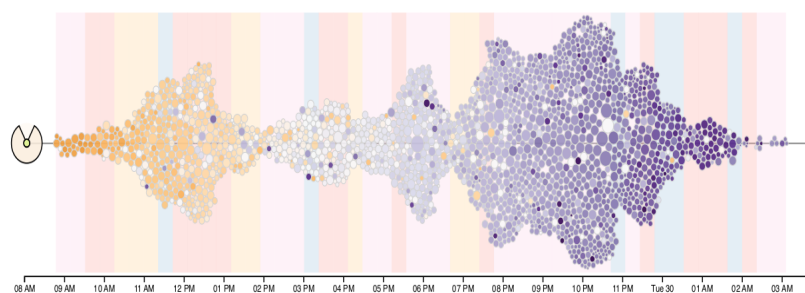


Figura 2.17: Diagrama de nó-link criado por Zhao et al. [48] para representar a difusão de informações em mídias sociais. Nesta visualização temos os 100 retweets anômalos classificados pelo modelo criado, durante o furacão Sandy em 2012. Os círculos indicam usuários participantes do Twitter nos tópicos, e as cores de fundo representam as variáveis de estado ocultas geradas pelo modelo, implicando as nuances dos padrões de divulgação de informações. Geralmente, durante este período de 18 horas, os números de anomalia dos usuários mudam de baixo (marrom) para alto (roxo), e há três grandes picos nos volumes de usuários na visualização apresentada, um ao meio-dia, outro às 18h e o último às 22h

um exemplo desse tipo de diagrama. O *Treemap* é um método de preenchimento de espaço que divide o espaço visual em regiões para mostrar relacionamentos hierárquicos. Com base no *Treemap* clássico, muitas extensões, como o *Treemap Voronoi* e o *Nmap*, se aplicam a diferentes tarefas analíticas e cenários de aplicação. A Figura 2.18 apresenta um *Nmap* criado por Duarte et al. [11].



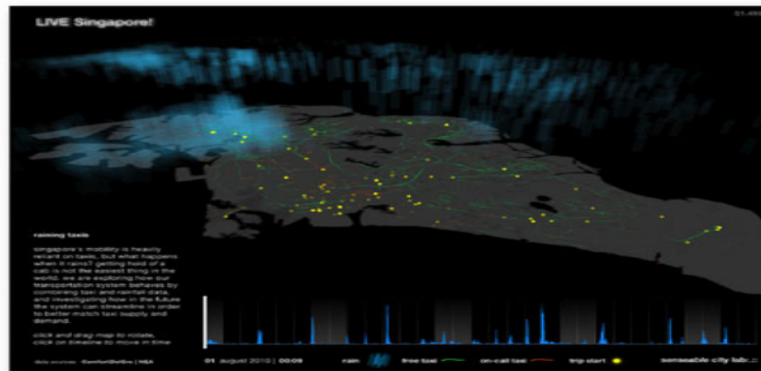
Figura 2.18: *Nmap* feito por Duarte et al. [11]. O *Nmap* representa dez anos de artigos científicas da conferência IEEE Infovis e alguns outros artigos frequentemente citados em tal conferência. Cada retângulo representa um grupo de documentos com conteúdo relacionado. Grupos similares são colocados próximos um do outro.

### 2.2.3 Visualização e abordagens de análise automatizada

Técnicas automatizadas de análise de dados permitem uma exploração conveniente de dados e descoberta de padrões, como agrupamentos, tendências e anomalias. Entretanto, algumas informações são difíceis de ser quantificadas com precisão em certas aplicações (por exemplo, preferências pessoais), o que dificulta o uso de abordagens automatizadas. Por outro lado, a visualização tornou-se uma técnica padrão para permitir o uso das capacidades perceptivas humanas no processo automatizado de exploração de dados. Através da combinação de visualização e abordagens analíticas automatizadas, tarefas de alto nível e complexas podem ser realizadas de forma mais eficaz e eficiente.

Nesta seção, discutimos como a visualização pode ser combinada com abordagens analíticas automatizadas, que é uma das principais questões da análise visual urbana. Nos últimos anos foram desenvolvidas um grande número de ferramentas que combinam a visualização com abordagens analíticas automatizadas, permitindo a exploração de dados e a interpretação de padrões no contexto urbano.

Kloeckl et al. [19] apresentam uma plataforma aberta para a coleta, combinação e distribuição de fluxos de dados urbanos em grande escala em tempo real, chamada *LIVE Singapore!*. Uma característica interessante do trabalho reside na correlação e integração de fontes de dados heterogêneas para uma visão mais ampla da dinâmica de uma cidade. No entanto, os autores utilizaram principalmente fluxos de dados de sensores e registros de detalhes de chamada. Dados gerados por usuários em alguma mídia social não foram incluídos nos mecanismos de análise da plataforma. As Figuras 2.19



(a)

Figura 2.19: Visualização presente na plataforma *LIVE Singapore!*. São apresentadas viagens e demandas de táxi em Cingapura durante fortes chuvas



(b)

Figura 2.20: Visualização presente na plataforma *LIVE Singapore!*. Distorção da forma da cidade de Cingapura em um mapa para refletir o tempo que leva para viajar da origem selecionada para outras partes da ilha

e 2.20 apresentam duas visualizações presentes no *LIVE Singapore!*.

Psyllidis et al. [32] apresentam o *SocialGlass*, uma plataforma Web para análise e integração de dados urbanos heterogêneos. A plataforma aproveita o potencial combinado de diversas fontes de dados urbanos, incluindo sensores e fluxos de mídia social (Twitter, Instagram e Foursquare), registros públicos disponíveis, e recursos de repositórios de conhecimento. Usando ciência dos dados, *crowdsourcing* e sobretudo integração e análise semântica dos dados, a plataforma permite o mapeamento de informações geográficas, visualização de padrões de deslocamento das pessoas, verificação da popularidade de locais, das condições de tráfego, bem como opiniões e preferências dos cidadãos e visitantes sobre locais específicos de uma cidade. As Figuras 2.21 e 2.22 são exemplos da interface inicial e de uma visualização presente no *SocialGlass*.

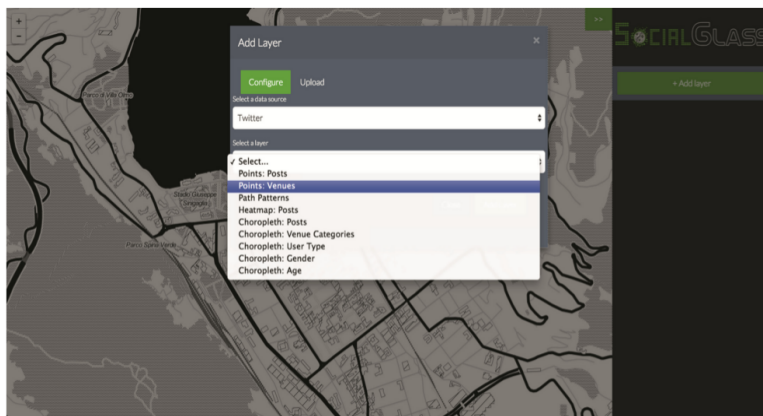


Figura 2.21: Menu inicial do *SocialGlass*, com várias opções de fontes e visualizações de dados



Figura 2.22: Agrupamentos de pontos dinâmicos de micro-postagens do Twitter, em combinação com padrões de caminhada de turistas estrangeiros e suas variações temporais na cidade de Roterdã

Buscando encontrar uma maneira automatizada de visualizar o ditado todos os caminhos levam a Roma, Reimann et al. [34] criaram mapas usando tanto algoritmos de roteamento que partem de múltiplos pontos de partida para um único ponto de destino, quanto algoritmos que partem de múltiplos pontos de destinos para um único ponto de partida. Utilizando a metodologia desenvolvida, os pesquisadores criaram diagramas de mobilidade para cidades selecionadas. Um desses diagramas foi batizado pelos autores de *Urban Mobility Fingerprint*, e é apresentado na Figura 2.23.

O diagrama evoluiu a partir da aplicação do princípio de roteamento de múltiplas origens para um único destino para uma área menor, no caso, uma cidade, nesse sentido, ele pode ser usado como uma ferramenta para analisar redes de mobilidade urbana. Na Figura 2.23 as cores são utilizadas para indicar a direção geográfica das rotas. O mapa



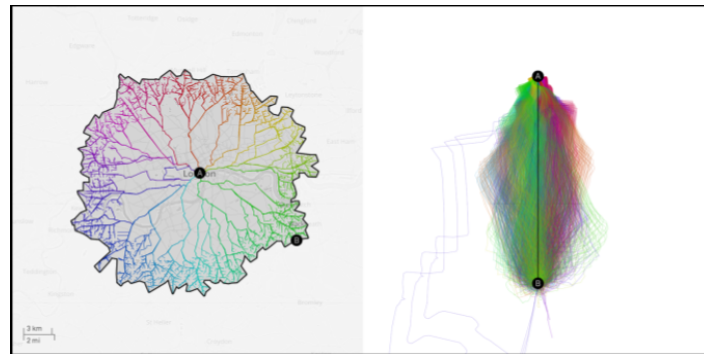


Figura 2.23: Exemplo do diagrama *Urban Mobility Fingerprint* criado por Reimann et al. [34] para a cidade de Londres.

do lado esquerdo do diagrama mostra os pontos alcançáveis em 15 minutos viajando de carro. A parte direita do diagrama exibe o *Urban Mobility Fingerprint*, no qual as cores também indicam a direção geográfica das rotas. Ele mostra o quanto cada rota desvia do caminho ótimo, que seria a conexão direta entre o ponto de origem e de destino. As rotas são redimensionadas para que tenham o mesmo tamanho. Quanto mais as rotas se desviam do caminho ótimo, mais indireta é a conexão. É importante ressaltar que os autores utilizam o tempo de viagem tanto para o cálculo das rotas, quanto para a determinação da conexão direta utilizada como referência.

Liu et al. [24] propuseram uma abordagem de análise visual para explorar a diversidade de rotas. O sistema suporta uma análise em várias escalas de diversidade que pode ajudar a revelar a importância das diferentes rotas para diferentes viagens e apoiar o planejamento e gerenciamento urbano. A Figura 2.24 apresenta uma visão geral desse sistema.

Zhong et al. [53] propõe um modelo de análise visual, que é aplicado a dados de transporte para gerenciar e extrair informações para estudos urbanos. Tal modelo é apresentado na Figura 2.25. O modelo pode ser dividido em três componentes:

- **Sistema de processamento de dados baseado em SIG (Sistema de Informação Geográfica):** Funciona como um mecanismo básico de processamento de dados coletados. A entrada é um conjunto de dados brutos. As saídas são visualizações de dados em diferentes níveis de detalhe. O primeiro nível produz conjuntos de dados limpos, o segundo produz dados agregados, gerados por simples consultas em bancos de dados, e o terceiro produz dados agregados gerados pelo módulo de análise de rede espaciais. Uma quarta visualização exibe o grafo da rede espacial, com os resultados gerados no módulo de análise de redes.
- **Módulo de análise de redes espaciais:** Nesse módulo, os autores utilizam

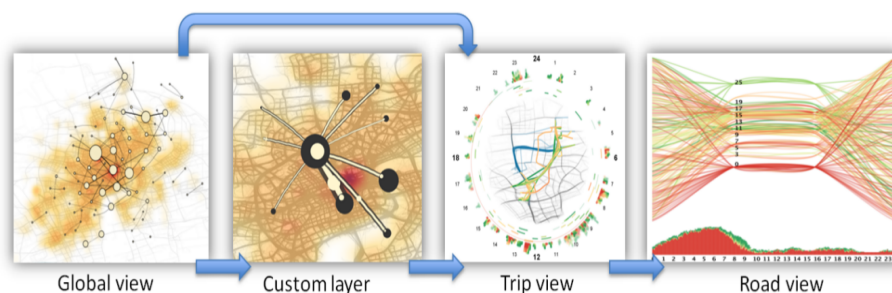


Figura 2.24: Visão geral do sistema criado por Liu et al. [24] para análise visual da diversidade de rotas. O sistema possui quatro componentes: 1) Visão global que mostra a diversidade geral de rotas de *hotspots* em uma cidade; 2) Camada personalizada que distorce a visão global para dar mais espaço a regiões com alta diversidade de rotas; 3) Visualização de viagem que exibe as trajetórias para um determinado par de origem e destino, bem como seus atributos espaciais e temporários associados; 4) Visualização de estrada, que apresenta informações estatísticas de todas as viagens que passam por uma determinada estrada

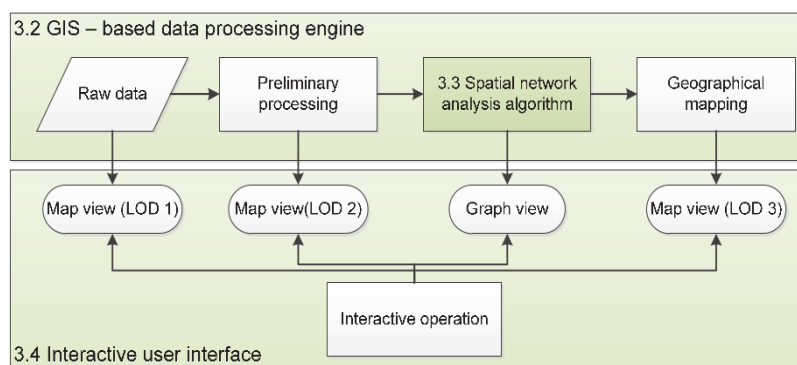


Figura 2.25: Modelo de análise visual, aplicado a dados de transporte em larga escala proposto por Zhong et al. [53]

dados de viagens dos indivíduos como uma aproximação para o processo de transferência de produtos e bens na cidade. Neste processo, um conjunto com milhões de registros de dados pode ser razoavelmente reduzido, através de agregação. O módulo também permite que índices de centralidade e organização espacial das áreas urbanas sejam extraídos a partir da análise dos padrões de movimentação.

- **Módulo de operações interativas:** O módulo consiste de uma interface gráfica que permite aos usuários selecionar e adicionar registros de viagem para realizar análises em tempo real.

Para validar e demonstrar o uso do modelo, os autores implementam uma ferramenta de visualização utilizando dados de transporte público da cidade de Singapura.

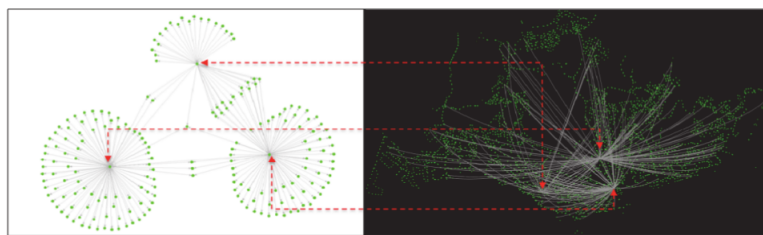


Figura 2.26: Duas visualizações da ferramenta criada por Zhong et al. [53], a primeira apresenta a visão de rede e a segunda a visão geográfica. Elementos em cada visualização são correspondentes uns aos outros. Este exemplo mostra as funções de ligação. Quando a ferramenta começa a carregar dados de viagens, a visualização geográfica adiciona ligações entre paradas do sistema de transporte público, e por sua vez, a visão de rede adiciona nós e arestas na rede.

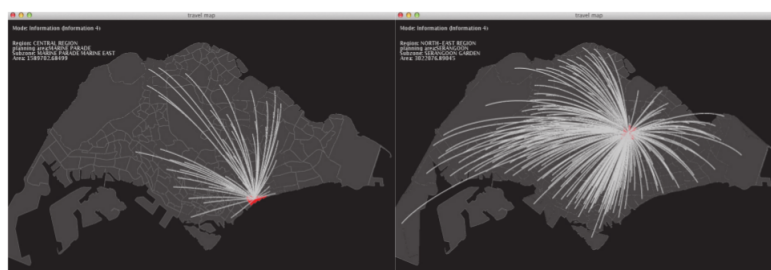


Figura 2.27: Visualização de fluxos no nível de subzona na ferramenta criada por Zhong et al. [53]. As viagens são agregadas por subzona. Ao selecionar uma zona individual, é possível obter informações detalhadas.

Os dados de entrada são um conjunto de amostras de dados de cartões de transporte público gerados pelo sistema de tarifas automáticas, e coletados em abril de 2011. Os eventos totalizam cerca de 5 milhões de registros de viagens.

A Figura 2.26 mostra os dois tipos de visualizações fornecidas pela ferramenta. A Figura 2.27 mostra informações de fluxos buscados e comparados na escala de subzona. Ao clicar em uma zona, todas as conexões entre essa zona e outras são mostradas como linhas curvas, como pode ser visto na Figura 2.27.

## 2.3 Considerações Finais

Os trabalhos apresentados neste capítulo mostram uma visão geral da área de Computação Urbana, com seus principais conceitos e projetos desenvolvidos, e que serviram como base para a construção do trabalho desenvolvido nessa dissertação. Os modelos propostos por Zheng et al. [49] e Zhong et al. [53] foram inspiração para a criação do modelo apresentado no Capítulo 3. Tal modelo generaliza o criado por Zhong et al.

[53], permitindo que seja o guia na criação de uma variedade maior de aplicações, e especializa o apresentado por Zheng et al. [49], focando em dados de transporte e trânsito. Dos tipos de dados apresentados, utilizamos sobretudo os dados geográficos, ou seja, as redes viária e de transporte, e os pontos de interesse. Fazemos uso das técnicas apresentadas na Seção 2.2.2 no projeto das visualizações criadas, focando nas técnicas de visualização baseadas em regiões e em linhas. Utilizamos as plataformas apresentadas na Seção 2.2.3 como base para a construção dos estudos de caso apresentados no Capítulo 4. Dessas plataformas, destaco a criada por Reimann et al. [34], cujo diagrama foi inspiração na construção da visualização apresentada no estudo de caso da Seção 4.1.

# Capítulo 3

## Metodologia

Neste capítulo é apresentado o modelo genérico para análise de dados urbanos, com foco em dados de transporte e trânsito, criado tendo como base as técnicas e métodos estudados. O modelo foi construído de forma genérica, permitindo que seja um guia na criação de aplicações e sistemas em Computação Urbana. Como pode ser visto na Figura 3.1, o modelo possui três camadas: **Integração**, **Armazenamento de dados** e **Prestação de serviços**. As camadas e seus respectivos componentes serão apresentadas e discutidas nas próximas seções.

### 3.1 Integração

Os dados produzidos em um ambiente urbano podem ser provenientes de diversas fontes, que podem ser tanto oficiais, ou seja, são dados produzidos por instituições governamentais (como a prefeitura ou órgão de trânsito de uma cidade), quanto não oficiais, ou seja, são criados por alguma empresa ou organização não necessariamente ligada ao governo de algum país (como o *OpenStreetMap*, um projeto de mapeamento colaborativo para criar um mapa livre e editável do mundo). Além disso, os dados urbanos podem ter natureza estática ou dinâmica. No caso de uma fonte de dados estática, os dados fornecidos por ela não mudam muito frequentemente e, dessa forma, não faz sentido que sua coleta seja feita em tempo real. Como exemplo temos a rede viária de uma cidade, que corresponde à estrutura de ruas e seus respectivos cruzamentos. Já no caso de uma fonte de dados dinâmica, os dados mudam frequentemente, e assim, sua coleta pode ser feita em tempo real. É o caso de dados coletados em redes sociais como o Twitter ou por sensores. Outra questão a ser ressaltada é que, como as fontes de dados são diversas e heterogêneas, os dados podem ser fornecidos em diversos formatos, desde formatos padronizados como o CSV (*Comma-Separated Values*),

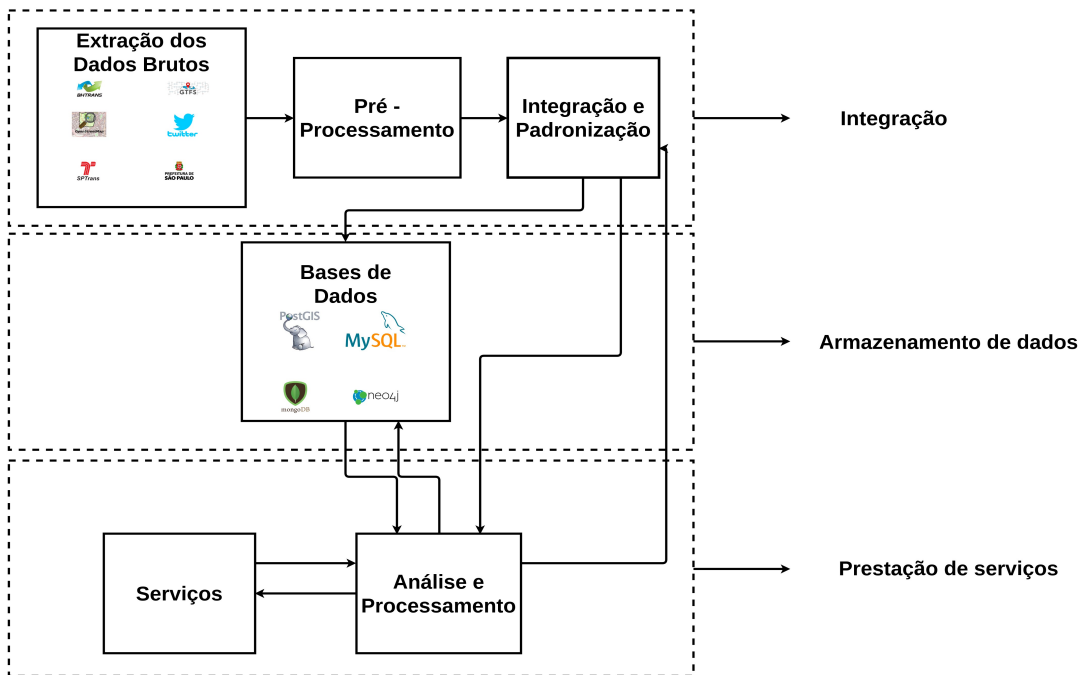


Figura 3.1: Modelo Geral para Computação Urbana. Foco em dados de transporte e trânsito

JSON (*JavaScript Object Notation*) e *Shapefile* (formato popular de arquivo contendo dados geoespaciais em forma de vetor), até conteúdo de texto livre, completamente não padronizado, como os de um *post* geolocalizado de uma rede social.

Diante desse cenário, as primeiras etapas para a construção de uma solução para análise de dados urbanos são a extração, a integração e a padronização dos dados. É justamente esse o papel desempenhado pela camada de Integração do modelo criado. Segundo a Figura 3.1, a camada de Integração possui três componentes: Extração dos dados brutos, Pré-processamento e Integração e Padronização, cada componente responsável parcial ou totalmente por alguma das etapas citadas.

### 3.1.1 Extração dos dados brutos

A componente de Extração dos dados brutos corresponde simplesmente ao processo de coleta dos dados a serem utilizados, seja através da obtenção de tais dados diretamente da fonte (forma comum de obtenção de dados estáticos), seja através do uso de alguma ferramenta fornecida pela fonte para obtenção dos dados, como exemplo, o uso de uma API (*Application Programming Interface*). Um exemplo de extração de dados é a obtenção das vias de circulação e interseções de vias de uma cidade a partir do *OpenStreetMap*, e do polígono delimitador de uma cidade a partir de um fonte oficial,

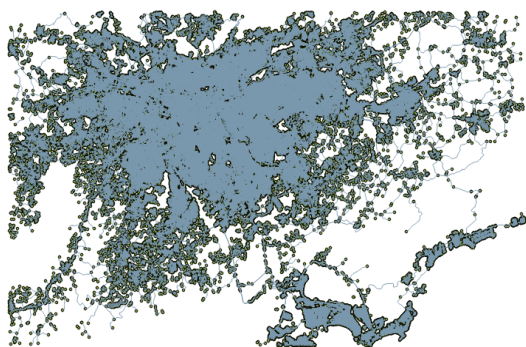


Figura 3.2: Rede viária completa da região metropolitana da cidade de São Paulo e cidades próximas extraída utilizando o *OpenStreetMap*.



Figura 3.3: Polígono que define os limites do Município da cidade de São Paulo.

como exemplo, o *website* da Prefeitura da cidade. A plataforma do *OpenStreetMap* permite, dentre outras formas, a extração dos dados contidos em sua base considerando uma região específica. Tal região é definida através de um retângulo delimitador que consiste de valores máximos e mínimos de latitude e longitude. Como exemplo, estão os dados da rede viária da região metropolitana da cidade de São Paulo e cidades próximas. Tais dados foram extraídos usando o retângulo limitador definido pelos valores de -47.1746 e -24.0102 para longitude e latitude mínimos, e -46.017 e -23.3538 para longitude e latitude máximos, o resultado é um tipo especial de arquivo no formato XML (*eXtensible Markup Language*), o OSM XML, que contém metadados específicos da plataforma. A Figura 3.3 apresenta o polígono que delimita os limites do Município da cidade de São Paulo, obtido no *website*<sup>1</sup> da Prefeitura da cidade.

### 3.1.2 Pré-Processamento

Dados brutos além de serem fornecidos em diferentes formatos, contêm inconsistências e não estão propriamente filtrados, ou seja, não contêm apenas os dados de interesse para a realização das análises desejadas. Dessa forma, o componente de Pré-Processamento usa algoritmos e ferramentas auxiliares de processamento para realizar a limpeza e a filtragem dos dados recém-extraídos de suas fontes.

Um exemplo em que a componente é utilizada para a filtragem de dados é na obtenção da rede viária de uma cidade a partir de dados brutos coletados no *OpenStreetMap*. Como dito, os dados fornecidos pela plataforma do *OpenStreetMap* são disponibilizados no formato OSM XML. Assim, para auxiliar na construção da base

<sup>1</sup>[http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/\\_SBC.aspx](http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx)

digital topológica da rede viária, utilizamos o software *osm2po*<sup>2</sup> que possui um *parser* de arquivos no formato OSM XML e auxilia na construção de dados roteáveis extraídos de arquivos nesse formato. O *parser* analisa o arquivo OSM XML e produz (dentre outros) arquivos de texto no formato CSV com os dados das arestas (vias) e dos nós (interseções de vias) da rede viária. Utilizando uma ferramenta auxiliar de processamento como o QGIS<sup>3</sup>, uma multiplataforma de georreferenciamento que provê visualização, edição e análise de dados georreferenciados, podemos visualizar a rede viária completa da região metropolitana da cidade de São Paulo e cidades próximas. Tal visualização é apresentada na Figura 3.2.

Como os dados apresentados nas Figuras 3.2 e 3.3 estão posicionados geograficamente no mesmo local do espaço, podemos selecionar com o auxílio do QGIS os nós e as vias da Figura 3.2 que estão contidos, ou seja, são completamente envolvidos pelo polígono da Figura 3.3. O resultado dessa filtragem é a rede viária do município da cidade de São Paulo, apresentada na Figura 3.4.

No processo descrito anteriormente, existe o uso concomitante de duas fontes de dados bem distintas. A primeira é o *OpenStreetMap*, uma fonte de dados não oficial, na qual foram coletados dados estáticos, em um formato específico, no caso, o formato OSM XML. A segunda é a Prefeitura de São Paulo, uma fonte de dados oficial, da qual também foram extraídos dados estáticos, porém no formato Shapefile. Dessa forma, o processo de filtragem e processamento para obtenção da rede viária do município da cidade de São Paulo pode ser visto como uma Pré-Integração de dados de fontes heterogêneas, preparando os dados brutos, recém-extraídos de suas fontes, para o processo de integração e padronização.

### 3.1.3 Integração e Padronização

Ao final do processo de Pré-Processamento, os dados já sofreram uma série de alterações visando sobretudo a sua filtragem e a remoção de inconsistências. Tal processo exige em alguns casos a combinação de fontes distintas para a seleção do conjunto de dados desejado, como foi descrito na seção anterior, processo chamado de Pré-Integração, para diferenciá-lo do processo de Integração realizado na componente de Integração e Padronização, apresentada nessa seção. Outra importante consequência do processo de pré-processamento é a uniformização do formato dos dados, ou seja, espera-se que após tal processo os dados estejam em um formato comum que facilite sua manipulação e integração. Para serem utilizados em conjunto os dados devem ser integrados,

---

<sup>2</sup><http://osm2po.de/>

<sup>3</sup><https://www.qgis.org>



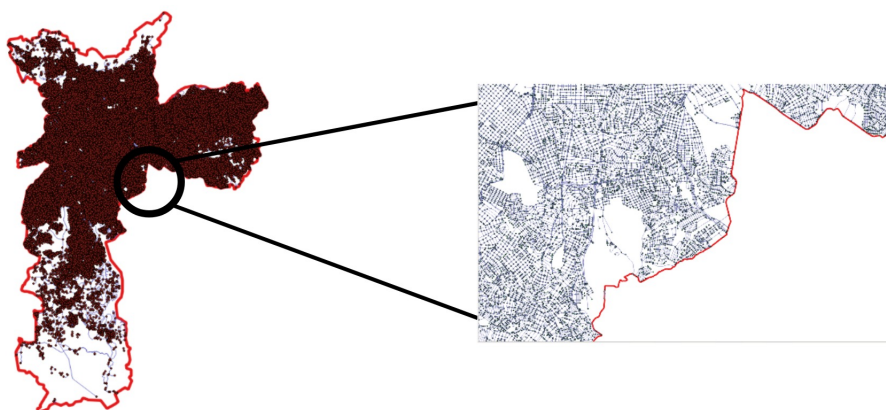


Figura 3.4: Rede Viária completa do Município da cidade de São Paulo. Cada ponto é um cruzamento de via. As vias foram subdivididas de modo a ligar os cruzamentos.

padronizados, e se necessário, armazenados em estruturas que possam ser facilmente acessadas pela camada de Prestação de serviços.

A Integração de fontes de dados consiste no uso de algoritmos e estruturas de armazenamento que permitam o uso simultâneo das informações provenientes dessas fontes. Essas estruturas podem ser tanto novos dados obtidos a partir da união das fontes, quanto estruturas que permitam a conexão dessas fontes, viabilizando seu uso conjunto. Um exemplo de Integração de fontes de dados para a produção de novos dados é a construção da rede de itinerários de ônibus de uma cidade a partir da união dos dados da rede viária da cidade com os dados de transporte público. Se os dados de transporte público têm a informação da localização geográfica das estações e a sequência de estações que cada meio de transporte percorre em seus trajetos, então podemos construir os itinerários como linhas que unam essas estações, seguindo a geometria descrita pela rede viária da cidade. Essa ideia é apresentada na Figura 3.5. O algoritmo criado para realizar tal integração, e responsável pela criação da rede de itinerários de ônibus da cidade de Belo Horizonte, é apresentado na Seção 4.1.1.

Na Padronização, os dados já pré-processados e integrados são estruturados de forma a seguir o padrão de nomenclatura e tipagem esperados pelos algoritmos e técnicas de análise implementados. Como exemplo, a aplicação ou sistema sendo criado pode estabelecer que: dados de redes (como redes viárias ou redes de itinerários de transporte público) sejam divididos em estruturas separadas, uma com informações dos nós, e outra com as informações das arestas, para cada nó da rede sejam informados um identificador do tipo inteiro, um campo textual para armazenagem do nome do nó, e um campo que informe a geometria do nó, e para cada aresta da rede seja informado um identificador do tipo inteiro, um campo textual para armazenagem do nome da aresta, dois campos inteiros que identifiquem os nós de origem e destino da

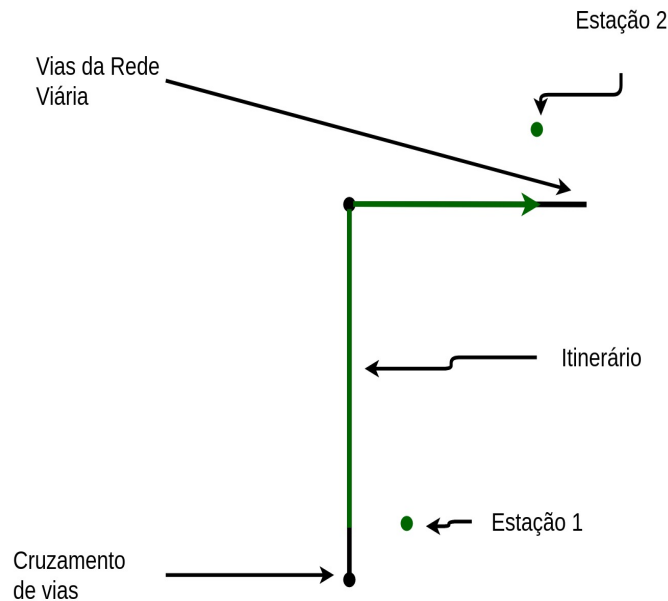


Figura 3.5: Integração entre os dados da rede viária de uma cidade e dados de transporte público. Se nos dados de transporte público temos que as estações 1 e 2 estão localizadas próximas às vias indicadas, e se a sequência do trajeto percorrido é da estação 1 para a estação 2, então, a ideia do algoritmo de integração é que o itinerário siga a geometria da rota que liga as estações 1 e 2 na rede viária.

aresta, e um campo que informe a geometria da aresta.

Após serem padronizados os dados podem ser persistidos na camada de Armazenamento de dados, ou ser enviados diretamente para processamento na camada de Prestação de Serviços, especificamente, para o componente de Análise e Processamento. Dados provenientes de fontes de dados dinâmicas, sobretudo as que são obtidas em tempo real, são em geral melhores candidatos para serem enviados diretamente para a camada de Prestação de Serviços, uma vez que, muitas vezes não é necessário persistir tais dados. O interesse é de fato no seu uso direto pelos algoritmos do componente de Análise e Processamento.

## 3.2 Armazenamento de dados

A forma mais comum para acessar, persistir, manipular e organizar dados de forma eficiente em sistemas de computação é através do uso de um Sistema de Gerenciamento de Bancos de Dados (SGBD). Assim, a camada de Armazenamento de dados no modelo proposto tem apenas um componente, Bases de dados, que corresponde ao(s)

Banco(s) de dados do(s) SGBD(s) utilizados na construção do sistema de análise de dados urbanos.

São vários os SGBDs que podem ser utilizados no contexto da análise de dados urbanos. Assim, a escolha de qual utilizar depende de uma série de fatores como tipo de dado a ser armazenado, se os dados são estruturados ou não, volume de dados, velocidade com a qual os dados serão inseridos no SGBD, funcionalidades que a aplicação necessita, etc. Dessa forma, determinar o SGBD mais adequado para o armazenamento dos dados utilizados por uma aplicação ou sistema é uma tarefa desafiante e que está fora do escopo desse trabalho. No entanto, como os dados urbanos são essencialmente georreferenciados, restringimos o escopo de SGBDs considerados no modelo aos SGBDs com extensões espaciais, como o PostGIS<sup>4</sup>, o MySQL Spatial<sup>5</sup>, o Neo4J Spatial<sup>6</sup>, MongoDB<sup>7</sup>, dentre outros. Considerando que as aplicações em Computação Urbana lidam com fontes de dados heterogêneas, e que essencialmente os dados produzidos por essas fontes, mesmo após serem pré-processados, integrados e padronizados tem características diferentes, considera-se que, em Computação Urbana, as soluções de armazenamento de dados utilizadas devem ser também heterogêneas, utilizando portanto um conjunto de SGBDs que atenda às necessidades do aplicação ou sistema projetado.

### 3.3 Prestação de serviços

A camada de Prestação de Serviços é composta por dois componentes: Análise e Processamento, e Serviço. Ela é a parte final de toda a cadeia de processamento criada no modelo apresentado na Figura 3.1. Nela é feita a entrega de algum serviço ou resultado para o usuário da aplicação ou sistema, seja retornando o resultado de um modelo de previsão implementado por algum algoritmo de mineração de dados ou aprendizado de máquina, seja apresentando alguma visualização construída, seja fazendo ambos. A ideia é que o resultado final produzido nessa camada tenha papel crucial nas tomadas de decisão sobre o planejamento urbano na cidade cujos dados foram analisados.

#### 3.3.1 Análise e Processamento

Ainda na Figura 3.1, o componente de Análise e Processamento comunica diretamente com o componente de Bases de Dados da camada de Armazenamento de dados e com o

---

<sup>4</sup><https://postgis.net/>

<sup>5</sup><https://dev.mysql.com/doc/refman/5.7/en/spatial-types.html>

<sup>6</sup><https://neo4j-contrib.github.io/spatial/>

<sup>7</sup><https://www.mongodb.com/>

componente de Integração e Padronização da camada de Integração. Isso ocorre porque ele consome os dados já integrados e padronizados provenientes de tais componentes.

O componente de Análise e Processamento é essencialmente a implementação do(s) algoritmo(s) de visualização de dados, mineração de dados e aprendizado de máquina que foram projetados para uso na aplicação ou sistema criado. Dessa forma, a camada espera que os dados retornados pelos componentes com as quais interfaceia estejam no padrão definido, garantindo assim a interoperabilidade das camadas do modelo.

### 3.3.2 Serviços

A componente de Serviços se comunica apenas com a componente de Análise e Processamento, como pode ser observado na Figura 3.1. Sendo assim, ela consome diretamente os resultados dos algoritmos de mineração de dados, aprendizado de máquina e visualização de dados, sendo responsável pela apresentação de tais resultados para o usuário final.

A construção dos elementos pertencentes à camada de Serviços deve ser feita de modo a apresentar os resultados provenientes da camada de Análise e Processamento de forma intuitiva e de fácil entendimento para o público-alvo da aplicação ou sistema criado. Assim, o componente é de suma importância para que o objetivo final do sistema de análise de dados urbanos criado seja atingido, que é o de auxiliar nas tomadas de decisão sobre o planejamento urbano, num processo recorrente para uma melhoria contínua e não obstrutiva da vida das pessoas, dos sistemas de funcionamento da cidade e do ambiente.

## 3.4 Considerações Finais

Como dito na Seção 2.3, o modelo criado teve como inspiração os modelos feitos por Zheng et al. [49] e Zhong et al. [53]. Ao ser comparado com o modelo criado por Zhong et al. [53], o modelo exposto neste capítulo tem em comum a ideia de obtenção dos dados brutos e de pré-processamento dos dados coletados. No entanto, o modelo de Zhong et al. [53] torna muito específico o processo de análise dos dados, possuindo uma camada na qual são utilizados apenas algoritmos de análise de redes espaciais, além de definir uma camada de interface com o usuário que utiliza técnicas de visualização específicas para dados no formato de rede.

Podemos estabelecer correlações entre as camadas do modelo de Zheng et al. [49] e as do modelo criado. A camada de *Urban Sensing and Data Acquisition* equivale

à camada de Integração, no entanto, a camada de Integração no modelo criado é dividida em componentes cujas funções vão além das especificadas por Zheng et al. [49] para a camada de *Urban Sensing and Data Acquisition*, que se restringe apenas à obtenção dos dados. No modelo de Zheng et al. [49] é definida uma camada de persistência de dados, equivalente à camada de Armazenamento de dados do modelo criado e além disso, Zheng et al. [49] também especificam a camada de *Urban Data Management* que é responsável pela criação de estruturas de indexação que permitem a consulta eficiente dos dados persistidos. Tal funcionalidade fica a cargo da camada de Armazenamento de dados no modelo apresentado, já que, em geral a criação de estruturas para consulta eficiente de dados é responsabilidade do sistema de gerência de dados utilizado. Finalmente, as camadas de *Urban Data Analytics* e *Service Providing* do modelo criado por Zheng et al. [49], equivalem respectivamente às componentes de Análise e Processamento, e Serviços da camada de Prestação de Serviços do modelo proposto. A grande diferença é na semântica utilizada por Zheng et al. [49] para a camada de *Service Providing*, que representa nesse modelo, o resultado final do uso do sistema ou aplicação desenvolvido, ou seja, a melhoria no planejamento urbano, a redução dos congestionamentos, etc. Já no modelo apresentado nesse capítulo, a componente de Serviços corresponde ao serviço final que interage diretamente com o usuário, ou seja, ferramentas de visualização de dados, ou resultados de algoritmos de análise automatizada, que se forem utilizados por um planejador urbano, podem culminar nos resultados previstos pela camada de *Service Providing*.

Assim, o modelo apresentado nesse capítulo generaliza o desenvolvido por Zhong et al. [53], permitindo que seja aplicado na construção de uma gama maior de aplicações, não apenas aquelas que utilizam dados de transporte e trânsito. Além disso, ele simplifica o modelo de Zheng et al. [49], criando camadas de processamento divididas em componentes com propósito específico, o que facilita o seu uso como guia na construção de aplicações em Computação Urbana.



# Capítulo 4

## Estudos de Caso

No presente capítulo são apresentados dois estudos de caso que utilizam o modelo apresentado e discutido no Capítulo 3. Para tal, mostramos como a própria construção de cada estudo de caso faz uso das camadas e componentes do modelo criado, comprovando assim a viabilidade, aplicabilidade e validade de tal modelo.

### 4.1 Estudo de caso I: *Impressão Digital de Trânsito Urbano*

Inspirados no trabalho de Reimann et al. [34], e com o intuito de aplicá-lo a dados de transporte e trânsito, criamos a *Impressão Digital de Trânsito Urbano*. A principal ideia que motivou a construção dessa visualização é permitir a comparação do acesso e da mobilidade urbana em uma cidade utilizando duas modalidades diferentes de transporte: o transporte público (mais especificamente os ônibus) e o transporte privado. A visualização se baseia na construção de rotas, e para tal, é necessário o acesso a dados roteáveis; nesse caso específico, aos dados da rede de circulação ou rede viária da cidade e aos dados da rede de itinerários de ônibus da cidade. Construímos a *Impressão Digital de Trânsito Urbano* utilizando os dados de duas metrópoles brasileiras: São Paulo e Belo Horizonte. Nas seções a seguir, apresentamos como a *Impressão Digital de Trânsito Urbano* foi construída utilizando o modelo proposto no Capítulo 3.

#### 4.1.1 Integração

Como dito, para a construir a *Impressão Digital de Trânsito Urbano* necessitamos dos dados da rede viária e da rede de itinerários de ônibus da cidade a ser analisada. Nas seções a seguir apresentamos como os dados de cada uma dessas redes, para as cidades

analisadas, foram extraídos de suas fontes, pré-processados, integrados e padronizados, de modo a serem armazenados e posteriormente utilizados na construção da visualização.

#### 4.1.1.1 Construção da base digital topológica da rede viária

Para a construção da base digital topológica da rede viária de uma cidade, necessitamos essencialmente de dois conjuntos de dados: as vias de circulação e os pontos de interseção dessas vias. Como discutido na Seção 3.1.1, utilizamos a plataforma do *OpenStreetMap* para a obtenção dos dados brutos das vias de circulação e das interseções de vias de uma cidade. Uma das formas de extrair dados de uma região nessa plataforma é através da definição de um retângulo delimitador. O principal problema dessa forma de extração é que são obtidos dados de vias e interseções de vias não apenas da cidade, mas também de cidades vizinhas. Para contornar esse problema, obtivemos também, de alguma fonte oficial como o *website* da Prefeitura da cidade, o polígono que define os limites do município da cidade a ser analisada. Dessa forma, podemos filtrar os dados obtidos do *OpenStreetMap*, nos concentrando apenas naqueles referentes à cidade em questão.

Os dados das vias de circulação e interseções de vias da Região Metropolitana de Belo Horizonte e cidades limítrofes foram extraídos da plataforma do *OpenStreetMap* utilizando como o retângulo limitador definido pelos valores de -44.1938 e -20.0392 para longitude e latitude mínimos, e -43.7318 e -19.7702 para longitude e latitude máximos. Utilizando um procedimento similar ao descrito na Seção 3.1.2, o *parser* presente no software *osm2po* foi usado para obter as arestas (vias) e os nós (interseções de vias) presentes no arquivo OSM XML extraído do *OpenStreetMap*. O resultado é apresentado na Figura 4.1. O polígono que delimita o município de Belo Horizonte, apresentado na Figura 4.2, foi obtido através de um serviço WFS (*Web Feature Service*)<sup>1</sup> oferecido pela Prefeitura de Belo Horizonte. O WFS é um dos serviços especificados pela OGC (*Open Geospatial Consortium*) para acesso e manipulação de dados geográficos na *web*, o serviço permite o acesso a dados, independentemente do formato de armazenamento. Existem vários tipos de softwares que podem ser clientes deste tipo de serviço. Neste trabalho utilizamos o QGIS, uma vez que o serviço retorna os dados no formato de arquivos Shapefile.

O próximo passo na construção da base digital topológica da rede viária é filtrar os dados presentes na Figura 4.1 que estão completamente contidos no polígono presente na Figura 4.2. De modo análogo ao descrito na Seção 3.1.2, tal filtragem foi feita

---

<sup>1</sup><http://bhmapogcbase.pbh.gov.br/bhmapogcbase/wfs>



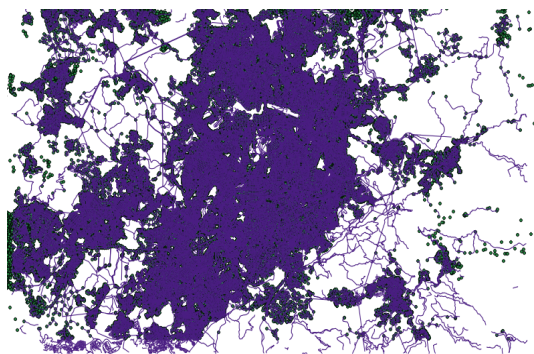


Figura 4.1: Rede viária completa da região metropolitana da cidade de Belo Horizonte e cidades próximas.



Figura 4.2: Polígono delimitador do Município da cidade de Belo Horizonte.

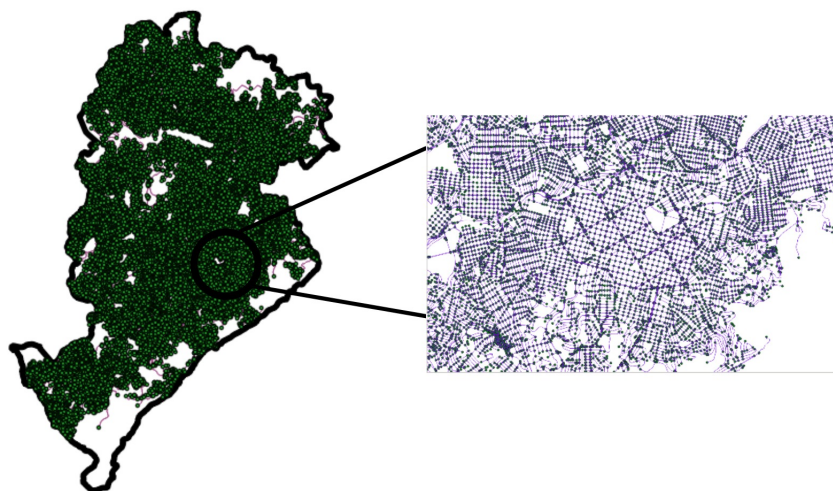


Figura 4.3: Rede Viária completa do município de Belo Horizonte, construída de modo análogo ao feito para a rede viária do município de São Paulo, presente na Figura 3.4

utilizando o software QGIS. O resultado é a rede viária do município da cidade de Belo Horizonte, apresentada na Figura 4.3.

Considerando a rede viária de uma cidade como um grafo, para que a *Impressão Digital de Trânsito Urbano* funcione de maneira similar ao *Urban Mobility Fingerprint* criado por Reimann et al. [34], precisamos que tal grafo tenha como valor de peso de cada aresta o tempo gasto para percorrê-la. No entanto, os dados presentes nas Figuras 4.3 e 3.4 não contêm tal informação. Outra questão a ser analisada é o fato de que os algoritmos de roteamento devem produzir rotas que respeitem as regras de trânsito, e dessa forma, os dados devem refletir tanto as situações nas quais as vias têm mão dupla, quanto aquelas nas quais elas têm mão única. Dito isso, os dados de ambas as redes viárias necessitam passar por um pré-processamento.

Estudos realizados pelos órgãos responsáveis pelo transporte público das metrô-

poles estudadas, a saber BHTrans<sup>2</sup> para a cidade de Belo Horizonte e SPTrans<sup>3</sup> para São Paulo, mostram dados sobre a velocidade média e tempo de viagem, tanto para o sistema de transporte público quanto para o privado. Para a cidade de Belo Horizonte, a velocidade média geral de circulação de um veículo particular é de 26,2 km/h, já para a cidade de São Paulo, o valor é de 18,9 km/h.

Por motivos de simplificação, no presente trabalho calculamos o tempo gasto para percorrer uma via de circulação considerando um veículo se locomovendo com os valores de velocidade média apresentados anteriormente. Essa é uma simplificação razoável, uma vez que, introduzir fatores extras como congestionamentos estão além do escopo deste trabalho. Assim, com o comprimento da via e a velocidade média de um veículo particular percorrendo tal via, podemos calcular o tempo de deslocamento desse veículo em segundos utilizando a Equação 4.1. Os dados extraídos do *OpenStreetMap* contêm uma coluna específica que auxilia nesse cálculo, a coluna *length*, que corresponde ao comprimento da via em metros.

$$\text{weight (s)} = \frac{((\text{length (m)}/1000) * 3600)}{\text{velocidade média(km/h)}} \quad (4.1)$$

Para que os algoritmos de roteamento criem rotas que respeitem as regras de trânsito, utilizamos os dados da coluna *oneway* dos dados, que indica se a via é de mão dupla ou de mão única. Caso o valor da coluna *oneway* seja falso, a via é de mão dupla. Assim, devemos criar uma nova entrada nos dados das vias da rede viária, invertendo os nós de origem e destino, e invertendo também a geometria da aresta, de modo que geograficamente ela esteja orientada no sentido contrário. Isso conclui o processo de pré-processamento dos dados das redes viárias.

O passo final para a conclusão da integração dos dados das redes viárias é a padronização dos mesmos, selecionando dentre as diversas colunas aquelas que são de interesse para o processo de análise, e que portanto devem ser persistidas na camada de armazenamento de dados. Como discutido na Seção 3.1.3, padronizamos os dados das redes viárias da seguinte forma:

1. Os dados são armazenados em dois arquivos, um com os dados dos nós, e outro com os dados das arestas da rede.
2. Para cada nó da rede existe: uma coluna com um identificador do tipo inteiro, uma coluna com o nome do nó, criado a partir do nome das vias que ele une, e

---

<sup>2</sup><http://www.bhtrans.pbh.gov.br/portal/page/portal/portalpublico/Temas/ObservatorioMobilidade>

<sup>3</sup>[https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/5reuniao\\_1389799252.pdf](https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/5reuniao_1389799252.pdf)

um campo que informa a geometria do nó, que no caso é um ponto.

3. Para cada aresta da rede existe: uma coluna com o valor de um identificador do tipo inteiro, uma coluna com o nome da aresta, que é o próprio nome da via, dois campos inteiros que identificam os nós de origem e destino da aresta, e um campo que informa a geometria da aresta, que no caso é uma poligonal.

#### 4.1.1.2 Construção da rede de itinerários de ônibus

A principal fonte utilizada nesse trabalho para a obtenção de dados relativos ao sistema de transporte público de uma cidade são os *feeds* da Especificação Geral de Feeds de Transporte Público (GTFS)<sup>4</sup> do *OpenStreetMap*, que define um formato comum para horários de transportes públicos e informações geográficas associadas. Os *feeds* GTFS permitem que as agências de transporte público publiquem suas informações e que os desenvolvedores criem aplicativos que consumam esses dados com interoperabilidade. Um *feed* GTFS é composto de uma série de arquivos de texto no formato CSV recolhidos em um arquivo ZIP. Cada arquivo modela um aspecto específico das informações sobre o transporte público: paradas, trajetos, viagens e outros dados relativos a horário. A agência de transporte público pode produzir um *feed* GTFS para compartilhar suas informações sobre transporte público com desenvolvedores, que criam ferramentas que consomem tais *feeds* para incorporar informações de transporte público em seus aplicativos. A GTFS pode ser usada para oferecer recursos mais poderosos a planejadores de viagens, publicadores de tabelas e uma ampla variedade de aplicativos que usam informações sobre transporte público de alguma forma. Os arquivos essenciais para que uma fonte de dados forneça um *feed* GTFS válido são:

- *agency.txt*: Fornece informações sobre uma ou mais agências de transporte público que fornecem os dados presentes no *feed*.
- *routes.txt*: Contém dados dos trajetos de cada modalidade de transporte público. Um trajeto é um grupo de viagens exibidas aos passageiros como um único serviço.
- *trips.txt*: Informa sobre as viagens de cada trajeto. Uma viagem é uma sequência de duas ou mais paradas que ocorrem em um horário específico.
- *stop\_times.txt*: Define os horários de partida e chegada dos veículos em paradas específicas em cada viagem, ou seja, o trajeto seguido pelos veículos em cada viagem.

---

<sup>4</sup><https://developers.google.com/transit/gtfs/reference?hl=pt-br>

---

**Algoritmo 1** Algoritmo para extração dos dados das rotas presentes no arquivo *routes.txt*

---

**Entrada:**

- i: índice do inteiro identificador da rota nas tuplas do arquivo *routes.txt*.
- j: índice do nome da rota nas tuplas do arquivo *routes.txt*.
- k: índice do tipo da rota nas tuplas do arquivo *routes.txt*.

**Saída:**

**Rotas** : Dicionário do tipo chave-valor. As chaves são os identificadores das rotas, os valores são tuplas com o nome e o tipo da rota.

```

1: procedimento PROCESSA ARQUIVO ROTAS
2:   Rotas ← {}
3:   para cada linha em routes.txt faça
4:     id_rota ← linha[i]
5:     nome_rota ← linha[j]
6:     tipo_rota ← linha[k]
7:     Rotas[id_rota] ← (nome_rota, tipo_rota)
8:   fim para
9:   retorna Rotas
10: fim procedimento

```

---

- *stops.txt*: Identificam as paradas, ou seja, locais individuais em que os veículos pegam ou deixam passageiros.
- *calendar.txt*: Define datas para serviços que usam uma programação semanal. Especificam quando o serviço começa e termina, bem como os dias da semana em que o serviço está disponível.

Um arquivo não obrigatório no padrão GTFS, mas que auxilia muito na construção da rede de itinerários de ônibus a partir dos dados do *feed*, é o arquivo *shapes.txt*. Esse arquivo contém a sequência de vértices ao longo do itinerário de cada linha de transporte, o que facilita o trabalho de visualização das trajetórias e o cálculo de distância percorrida. O arquivo *stops.txt* já contém os dados necessários para a inclusão dos nós da rede de itinerários de ônibus. Os Algoritmos de 1 a 5 apresentam como é feito o pré-processamento dos arquivos *routes.txt*, *trips.txt*, *stop\_times.txt*, *stops.txt* e *shapes.txt*. Tais algoritmos são *parsers* de arquivos no formato CSV que apenas percorrem os arquivos linha a linha, extraindo os dados das colunas desejadas, e criando ao final dicionários do tipo chave-valor com os dados extraídos.

O Algoritmo 6 realiza a construção das arestas da rede de itinerários de ônibus de uma cidade a partir dos dados pré-processados pelos Algoritmos de 1 a 5. O objetivo principal do Algoritmo 6 é produzir uma lista de tuplas com informações sobre os

---

**Algoritmo 2** Algoritmo para extração dos dados das viagens presentes no arquivo *trips.txt*

---

**Entrada:**

- i: índice do inteiro identificador da rota nas tuplas do arquivo *trips.txt*.
- j: índice do inteiro identificador da viagem nas tuplas do arquivo *trips.txt*.
- k: índice do inteiro identificador do *shape* da viagem, m nas tuplas do arquivo *trips.txt*.

**Saída:**

**Viagens** : Dicionário do tipo chave-valor. As chaves são os identificadores das viagens, os valores são tuplas com os identificadores da rota e do *shape* de cada viagem.

```

1: procedimento PROCESSA ARQUIVO VIAGENS
2:   Viagens ← {}
3:   para cada linha em trips.txt faça
4:     id_rota ← linha[i]
5:     id_viagem ← linha[j]
6:     id_shape ← linha[k]
7:     se Rotas[id_rota].tipo_rota == 3 então      ▷ % Rotas do tipo 3 são de
                                                    ônibus segundo a especi-
                                                    ficação do GTFS %
8:       Viagens[id_viagem] ← (id_rota, id_shape)
9:     fim se
10:  fim para
11:  retorna Viagens
12: fim procedimento

```

---

trechos dos trajetos de cada viagem de ônibus. As informações são: o nome da rota a qual o trecho pertence, os identificadores dos pontos de paradas de origem e destino do trecho, o tempo em segundos que um veículo leva para ir da parada de origem para a de destino, e a geometria da linha que liga as paradas de origem e destino. Assim, a ideia do algoritmo é obter inicialmente para cada trajeto o nome da rota e o *shape* ao qual ele pertence, e em seguida percorrer a lista de tuplas que constituem o trajeto em pares, sendo cada par correspondente a um nó origem e destino de uma aresta. Para cada par são extraídos o horário de saída da origem e o horário de chegada no destino, e assim computado o tempo para ir de um nó ao outro. Além disso, para obtermos a geometria da linha que liga os nós, obtemos primeiramente os nós pertencentes ao *shape* do trajeto, o nó do *shape* do trajeto mais próximo da origem e o nó mais próximo do destino, determinamos então que a geometria da linha que liga o nó origem ao nó destino é a linha que liga os pontos do *shape* que estão entre o nó mais próximo da origem e o nó mais próximo do destino. Ao realizar esse processo para todos os trajetos, temos as tuplas com as informações necessárias para a construção das arestas da rede

---

**Algoritmo 3** Algoritmo para extração dos dados dos *shapes* presentes no arquivo *shapes.txt*

---

**Entrada:**

- i: índice do inteiro identificador do *shape* nas tuplas do arquivo *shapes.txt*.
- j: índice da latitude dos pontos do *shape* nas tuplas do arquivo *shapes.txt*.
- k: índice da longitude dos pontos do *shape* nas tuplas do arquivo *shapes.txt*.

**Saída:**

**Shapes** : Dicionário do tipo chave-valor. As chaves são os identificadores dos *shapes* das viagens, os valores são listas com os pontos pertencentes a cada *shape*.

```

1: procedimento PROCESSA ARQUIVO SHAPES
2:   Shapes ← {}
3:   para cada linha em shapes.txt faça
4:     id_shape ← linha[i]
5:     latitude_ponto ← linha[j]
6:     longitude_ponto ← linha[k]
7:     Shapes[id_shape] += cria_ponto(latitude_ponto, longitude_ponto)
8:   fim para
9:   retorna Shapes
10: fim procedimento

```

---

de itinerários de ônibus.

O *feed* GTFS para a cidade de São Paulo é fornecido pela agência de trânsito responsável pela cidade, ou seja, a SPTrans<sup>5</sup>. Os dados do arquivo *shapes.txt* para a cidade de São Paulo são apresentados na Figura 4.4. Ele contém dados de todas as modalidades de transporte cobertas pelo GTFS, incluindo o sistema de ônibus. Os dados produzidos após o uso dos Algoritmos de 1 a 6 nos arquivos do *feed* GTFS de São Paulo são apresentados na Figura 4.5. Pela própria construção dos dados (sobretudo pela filtragem feita no Algoritmo 2), é garantido que as arestas presentes na Figura 4.5 são apenas de trajetos de ônibus.

Os dados do *feed* GTFS da cidade de Belo Horizonte são fornecidos pela agência de trânsito BHTrans<sup>6</sup>. A Figura 4.6 apresenta os pontos que formam os *shapes* para o sistema de transporte público da cidade de Belo Horizonte. Como pode ser percebido, os dados fornecidos pelo *feed* são extremamente incompletos. Dessa forma, a construção da rede de itinerários de ônibus da cidade de Belo Horizonte não pode ser feita com o uso do Algoritmo 6.

Os arquivos *routes.txt*, *trips.txt*, *stop\_times.txt* e *stops.txt* fornecidos no *feed* de Belo Horizonte não têm problemas, e assim são pré-processados utilizando os Algorit-

---

<sup>5</sup><http://www.sptrans.com.br/desenvolvedores/GTFS.aspx>

<sup>6</sup><https://transitfeeds.com/p/bhtrans/640>

---

**Algoritmo 4** Algoritmo para extração dos dados dos pontos de parada presentes no arquivo *stops.txt*

---

**Entrada:**

- i: índice do inteiro identificador do ponto de parada nas tuplas do arquivo *stops.txt*.
- j: índice da latitude dos pontos de parada nas tuplas do arquivo *stops.txt*.
- k: índice da longitude dos pontos de parada nas tuplas do arquivo *stops.txt*.

**Saída:**

Paradas : Dicionário do tipo chave-valor. As chaves são os identificadores dos pontos de parada, os valores são a geometria dos pontos de parada.

```

1: procedimento PROCESSA ARQUIVO PARADAS
2:   Paradas  $\leftarrow$  {}
3:   para cada linha em stops.txt faça
4:     id_parada  $\leftarrow$  linha[i]
5:     latitude_ponto  $\leftarrow$  linha[j]
6:     longitude_ponto  $\leftarrow$  linha[k]
7:     Paradas[id_parada]  $\leftarrow$  cria_ponto(latitude_ponto, longitude_ponto)
8:   fim para
9:   retorna Paradas
10: fim procedimento

```

---



Figura 4.4: Pontos dos *shapes* presentes no arquivo *shapes.txt* para a cidade de São Paulo.

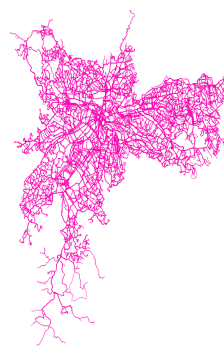


Figura 4.5: Arestas da rede de itinerários de ônibus da cidade de São Paulo.

mos 1, 2, 4, e 5. O fato do arquivo *shapes.txt* ser incompleto no *feed* de Belo Horizonte dificulta a obtenção da geometria dos trechos dos trajetos descritos no arquivo *stop\_times.txt*. Para solucionar esse problema, uma solução é aproximar a geometria das linhas de cada trecho de trajeto de ônibus pelo caminho mínimo que une os pontos de parada de origem e destino desses trechos na rede viária da cidade. Essa ideia é apresentada na Figura 3.5. Isso é uma aproximação razoável, uma vez que os ônibus trafegam de fato nos trechos de vias da rede viária, e os pontos de ônibus de origem e destino de um trecho de trajeto não estão em geral muito distantes um do outro,

---

**Algoritmo 5** Algoritmo para extração dos dados dos trajetos de cada viagem, presentes no arquivo *stop\_times.txt*

---

**Entrada:**

i: índice do inteiro identificador da viagem a qual cada trajeto está associado nas tuplas do arquivo *stop\_times.txt*.

j: índice do inteiro identificador dos pontos de parada em cada trajeto nas tuplas do arquivo *stop\_times.txt*.

k: índice do horário de chegada dos veículos na parada atual nas tuplas do arquivo *stop\_times.txt*.

l: índice do horário de saída dos veículos da parada atual nas tuplas do arquivo *stop\_times.txt*.

Viagens: dicionário com os dados das viagens contidas no arquivo *trips.txt* pré-processadas.

**Saída:**

Trajetos : Dicionário do tipo chave-valor. As chaves são os identificadores das viagens, os valores são listas que correspondem às sequências de paradas, e os horários de chegada e saída dos veículos que formam o trajeto da viagem.

```

1: procedimento PROCESSAR_ARQUIVO_TRAJETOS
2:   Trajetos ← {}
3:   para cada linha em stop_times.txt faça
4:     id_viagem ← linha[i]
5:     id_parada ← linha[j]
6:     horario_chegada ← linha[k]
7:     horario_saida ← linha[l]
8:     se id_viagem ∈ chaves do dicionário Viagens então
9:       Trajetos[id_viagem] += (horario_chegada, horario_saida, id_parada)
10:    fim se
11:  fim para
12:  retorna Trajetos
13: fim procedimento

```

---

o que torna a aproximação pelo caminho mínimo muito próxima da geometria efetiva do trecho do itinerário. Tal solução indica que deve ser feita uma integração entre os dados do *feed* GTFS e da rede viária de Belo Horizonte. A ideia dessa integração é construir uma **rede de circulação híbrida roteável**, na qual os nós podem ser tanto interseções de vias, quanto pontos de paradas de ônibus, e assim possamos obter o caminho mínimo entre dois pontos de parada de ônibus seguindo a geometria das vias da rede viária.

O primeiro passo para a construção dessa rede de circulação híbrida é localizar os pontos de paradas de ônibus presentes no arquivo *stops.txt* na via de circulação da rede viária mais próxima dele, ou seja, criar associações entre as vias da rede viária





Figura 4.6: Pontos que constituem os *shapes* dos trajetos do sistema de transporte público da cidade de Belo Horizonte.

e os pontos de parada que podem ser localizados nas mesmas. Tais associações são criadas pelo Algoritmo 7. Ele percorre todos os dados de pontos de parada (já pré-processados), e obtém para cada um a via da rede viária mais próxima dele e a fração segundo a qual o ponto correspondente à sua projeção na via divide a geometria da mesma. Assim, ao final da execução do Algoritmo 7, existe um dicionário informando, para cada via, a lista de pontos de parada localizados nas mesmas, bem como as frações segundo as quais tais pontos as dividem.

A rede de circulação híbrida é construída com o Algoritmo 8. A ideia desse algoritmo é utilizar as associações criadas pelo Algoritmo 7 para subdividir (se necessário) as vias da rede viária, criando subvias que ligam tanto as interseções de vias aos pontos de parada, quanto os pontos de parada entre si. Assim, para cada via da rede viária, o algoritmo inicialmente obtém todos os pontos de parada de ônibus que podem estar geograficamente localizados nela. Caso haja pontos de parada associados à via, ela deve ser subdividida. Para que as subdivisões sejam feitas corretamente, os vértices originais de origem e destino da via são adicionados à lista de pontos de parada, o vértice de origem é adicionado no início da lista (enfileirado), e o vértice de destino no final da lista (empilhado), sendo as frações correspondentes de tais vértices 0 e 1 respectivamente. Em seguida, a lista de pontos de parada é ordenada de acordo com os valores das frações de cada ponto. Finalmente, a lista de pontos de parada é percorrida em pares. Como essa lista está ordenada pelo valores das frações, em cada par esses valores definem os limites de cada subdivisão da via, ou seja, criam subvias. São criados então dois vértices na rede de circulação híbrida para cada par de pontos, e uma aresta (subvia) ligando cada par. Caso não haja pontos de parada associados à via analisada, ela é adicionada sem subdivisões na rede de circulação híbrida, ou seja, seus nós de origem e destino são adicionados à lista de vértices e a própria via é adicionada à lista de arestas.

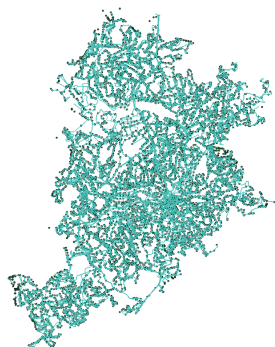


Figura 4.7: Rede de itinerários de ônibus da cidade de Belo Horizonte.

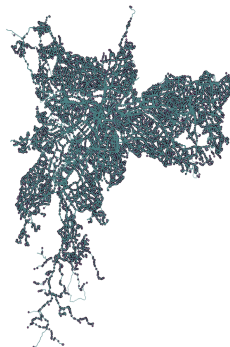


Figura 4.8: Rede de itinerários de ônibus da cidade de São Paulo.

Para a construção das arestas da rede de itinerários de ônibus de Belo Horizonte é utilizado o Algoritmo 9. Esse algoritmo é muito similar ao Algoritmo 6, com a diferença crucial que os trechos de cada trajeto dos itinerários de ônibus são obtidos através do cálculo do caminho mínimo na rede de circulação híbrida, ligando os pontos de parada de origem e destino de cada trecho. Esse caminho mínimo é obtido utilizando o Algoritmo de Dijkstra. Assim, o Algoritmo 9 percorre a lista de tuplas que constituem cada trajeto em pares, sendo cada par correspondente a um nó origem e um nó destino de uma aresta, e define que a geometria da linha que liga os nós é formada pelo conjunto de arestas que formam o caminho mínimo entre esses nós na rede de circulação híbrida. O algoritmo utiliza uma estratégia *best-effort*, na qual, caso o caminho mínimo entre os nós não possa ser encontrado, ele é aproximado por um segmento de reta que os liga.

Finalmente, os dados de nós e arestas das redes de itinerários de ônibus de Belo Horizonte e São Paulo são padronizados de modo análogo ao apresentado na Seção 4.1.1.1 para a construção dos dados das redes viárias. As Figuras 4.7 e 4.8 apresentam o resultado do processo de integração para a construção das redes dessas duas cidades.

#### 4.1.2 Armazenamento de dados

Ao final do processo de integração, são gerados dois arquivos no formato CSV com os dados dos nós e arestas de cada uma das redes analisadas, totalizando quatro arquivos por cidade. Como estamos analisando duas cidades, são oito arquivos no formato CSV. Como já discutimos, esses arquivos têm essencialmente a mesma estrutura, ou seja, todos os arquivos com dados de nós informam um identificador do tipo inteiro, o nome do nó, e a geometria do nó; e todos os arquivos com dados de arestas informam

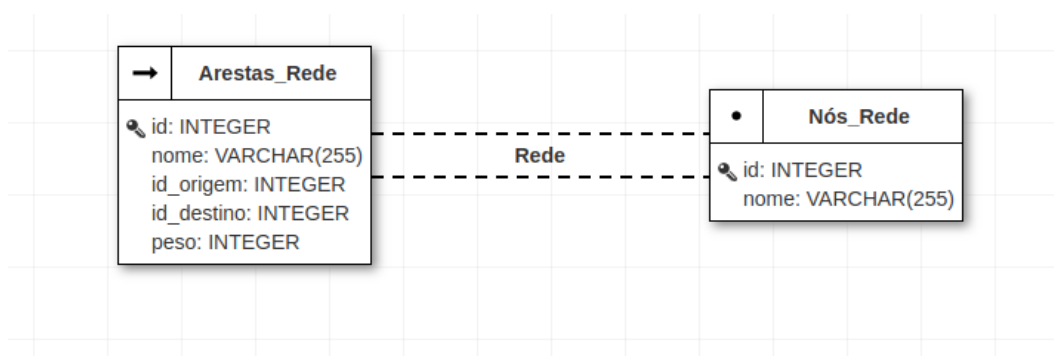


Figura 4.9: Diagrama que representa a estrutura geral e relacionamentos das tabelas utilizadas para persistência dos dados das redes viária e de itinerários de ônibus.

um identificador do tipo inteiro, o nome da aresta, dois campos inteiros que identificam os nós de origem e destino da aresta, e a geometria da aresta.

Como apresentado na Seção 3.2, o armazenamento persistente de dados é feito em geral utilizando um SGBD. São diversos os SGBDs disponíveis para uso. No entanto, como na construção da *Impressão Digital de Trânsito Urbano* lidamos com dados georreferenciados, nos restringimos àqueles que possuem extensões espaciais, como o PostGIS, o MySQL Spatial e o Neo4J Spatial. Dado o volume de dados tratados (em média 43180 nós e 146846 arestas), e a relação intrínseca entre eles (os dados dos nós são diretamente referenciados pelas arestas), escolhemos utilizar o PostGIS, pelo conhecimento prévio do autor tanto da estrutura quanto das funcionalidades oferecidas por esse sistema.

A Figura 4.9 mostra a estrutura geral das tabelas criadas no PostGIS para armazenagem dos dados das redes viária e de itinerários de ônibus. Os campos de cada tabela têm relação direta com as colunas dos arquivos CSV criados pelo processo de integração. Assim, a importação dos dados é trivial. É importante ressaltar o relacionamento *um-para-muitos* entre as tabelas de nós e as tabelas de arestas, indicando que um nó pode ser referenciado como nó origem ou destino de uma ou mais arestas. São criadas então um total oito tabelas, correspondentes aos oito arquivos CSV com dados de nós e arestas das redes. Uma vez persistidos, os dados das redes podem então ser facilmente acessados e manipulados pelos algoritmos desenvolvidos na camada de prestação de serviços.

### 4.1.3 Prestação de serviços

Para a construção da *Impressão Digital de Trânsito Urbano* é necessário o uso de algoritmos capazes de analisar e processar os dados das redes viária e de itinerários de

ônibus de uma cidade. Especificamente, tais algoritmos devem ser capazes de:

1. Determinar a partir de qualquer ponto da cidade os nós de determinada rede que são alcançáveis em um intervalo de tempo específico, e as rotas para tais nós.
2. Definir uma linha reta de referência ligando o ponto de origem selecionado a um dos pontos alcançáveis. Essa linha deve ser selecionada de modo a cobrir visualmente todas as rotas obtidas.
3. Rotacionar as rotas de modo que elas fiquem verticalmente alinhadas na direção de sul para norte, e redimensioná-las, de modo que todas elas aparentem iniciar no ponto de origem selecionado e terminem no ponto final da reta de referência.

Para desempenhar a primeira tarefa utilizamos uma extensão do SGBD PostGIS chamada *pgRouting*<sup>7</sup>, que fornece funcionalidades de roteamento geoespacial. Dentre os diversos algoritmos presentes na biblioteca do *pgRouting* destacamos:

- *Driving Distance*: Usando o algoritmo de Dijkstra, o algoritmo obtém todos os nós cujas rotas a partir de um nó de origem tem custo total menor ou igual que o especificado.
- *K-Dijkstra*: Retorna o caminho mínimo utilizando o algoritmo de Dijkstra entre um nó de origem e um conjunto de nós de destino.

Assim, dado um ponto qualquer da cidade, primeiramente devemos identificar qual o nó da rede selecionada (rede viária ou ponto de parada) mais próximo do ponto escolhido. Por construção, tanto as arestas das redes viárias quanto as das redes de itinerários de ônibus têm como valor de peso o tempo necessário para percorrê-las. Assim, de posse do nó da rede mais próximo do escolhido, podemos utilizar o algoritmo *Driving Distance* para obter os pontos da rede que são alcançáveis a partir desse nó em determinado intervalo de tempo. Finalmente, utilizando o algoritmo *K-Dijkstra*, obtemos as rotas entre o ponto selecionado e os pontos alcançáveis obtidos. Por questões de performance e principalmente para evitar a poluição visual na construção da visualização, optamos por limitar o número de nós alcançáveis (e portanto de rotas calculadas) aos 100 mais distantes em termos de tempo de viagem, obtendo portanto nós cuja distância da origem em termos de tempo é mais próxima do intervalo de tempo selecionado.

Para realizar a segunda tarefa, construímos a consulta que calcula as rotas de modo a ordená-las de forma decrescente pela distância entre o nó de origem e destino,

---

<sup>7</sup><https://pgrouting.org/>

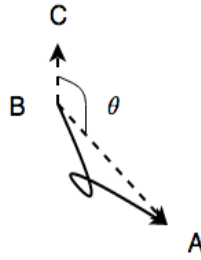


Figura 4.10: Determinação do ângulo de rotação que alinhe uma rota verticalmente na direção de sul para norte

além disso, retornamos também a linha reta que une tais pontos para cada rota. Assim, a linha de referência é simplesmente a linha reta que une os pontos de origem e destino da primeira rota retornada por tal consulta, assumimos então que a linha reta que "cobre" a maior distância é capaz também de "cobrir" visualmente todas as rotas obtidas.

Finalmente, para a terceira tarefa precisamos: definir o ângulo de rotação que faz com que as rotas e a linha de referência fiquem verticalmente alinhadas na direção de sul para norte, e calcular o fator de redimensionamento de cada rota. Dados três pontos A, B e C, podemos calcular os vetores  $\overrightarrow{AB}$  e  $\overrightarrow{BC}$  como:

$$\overrightarrow{AB} = B - A \quad \overrightarrow{BC} = C - B$$

Sendo  $\cdot$  o produto escalar entre vetores, e  $\| * \|$  o módulo de um vetor, o ângulo entre os pontos A, B e C é o ângulo entre os vetores  $\overrightarrow{AB}$  e  $\overrightarrow{BC}$ , calculado através da fórmula:

$$\theta = \arccos \left( \frac{\overrightarrow{AB} \cdot \overrightarrow{BC}}{\|\overrightarrow{AB}\| \|\overrightarrow{BC}\|} \right) \quad (4.2)$$

Assim, conforme ilustrado na Figura 4.10, o ângulo de rotação  $\theta$  para que uma rota fique alinhada verticalmente na direção sul-norte pode ser calculado como o ângulo entre o ponto final da rota (A), o ponto de origem da rota (B), e qualquer ponto com valor de longitude igual ao do ponto origem da rota e latitude maior que a desse ponto (C).

Com o valor do ângulo de rotação de cada rota definido, precisamos calcular o fator de redimensionamento das rotas. O objetivo do redimensionamento é fazer com que as rotas aparentem iniciar no ponto de origem selecionado e terminem no ponto final da reta de referência. Para tal, devemos aumentar ou diminuir as rotas de modo que o ponto final de todas elas coincidam com o ponto final da reta de referência. Como temos a distância em metros entre os pontos de origem e destino da reta de referência

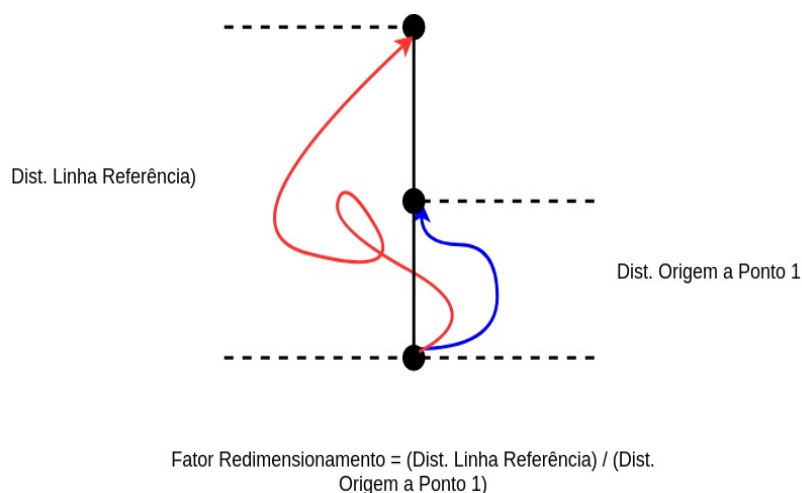


Figura 4.11: Cálculo do fator de redimensionamento de uma rota na construção da *Impressão Digital de Trânsito Urbano*.

e das rotas, o fator de redimensionamento de cada rota pode ser calculado utilizando a fórmula apresentada na Figura 4.11.

O *Urban Mobility Fingerprint* permite a comparação visual do quanto as rotas criadas se desviam do caminho ótimo (uma linha reta) entre os pontos de origem e destino. Com o objetivo de quantificar esse desvio criamos um indicador, chamado *Indicador de Ineficiência da Rota*, calculado como o quociente entre o tamanho em metros da rota e a distância euclidiana (também em metros) entre os pontos de origem e destino. A ideia é que, como a linha reta que une os pontos de origem e destino é a rota de menor tamanho possível entre esses pontos, o indicador quantifica o quão distante deste caminho ideal está a rota criada. Além de calcular esse indicador para cada rota, calculamos também a mediana dos indicadores e o tempo médio de viagem das rotas obtidas.

Os algoritmos anteriormente discutidos são disponibilizados na forma de *webservices*, cujos resultados são consumidos diretamente pelo **serviço** final, responsável pela criação de uma interface com o usuário para apresentação desses resultados. Tal serviço foi implementado como uma aplicação *web*. A Figura 4.12 apresenta a interface de consulta disponibilizada aos usuários para cada uma das cidades. Nela o usuário pode:

1. Selecionar inicialmente a rede na qual as rotas serão calculadas.
2. Decidir o ponto de origem das rotas, que pode ser tanto um ponto qualquer selecionado no mapa, quanto um dos pontos de interesse previamente cadastrados.

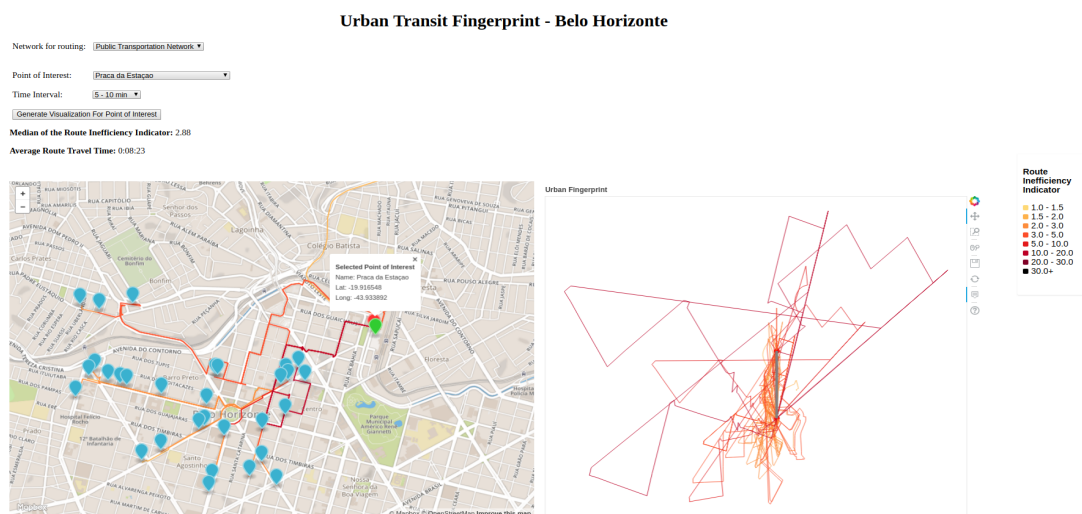


Figura 4.12: Resultados da aplicação da *Impressão Digital de Trânsito Urbano* para Belo Horizonte. A consulta foi feita pelos os pontos da rede de itinerários de ônibus de Belo Horizonte que são alcançáveis em um intervalo de 5 a 10 minutos, a partir de um ponto pré-selecionado da cidade, no caso, a Praça da Estação.

3. Selecionar o intervalo de tempo que deve ser considerado no algoritmo para obtenção dos pontos alcançáveis a partir da origem.

O sistema então calculará, utilizando os algoritmos anteriormente discutidos:

- Os 100 pontos da rede selecionada mais distantes em termos de tempo de viagem, que são alcançáveis a partir do ponto de origem, no intervalo de tempo informado, além das rotas para esses pontos.
- O *Indicador de Ineficiência da Rota* de cada rota, bem como a mediana desse indicador, e o tempo médio de viagem para todas as rotas obtidas. Nas visualizações, o valor do indicador de cada rota é denotado usando a cor da linha, definida utilizando a associação entre cores e faixas de valores apresentadas na legenda.
- A visualização da *Impressão Digital de Trânsito Urbano*, que apresenta as rotas rotacionadas e redimensionadas, e a linha reta de referência.

As visualizações produzidas pela aplicação são interativas, permitindo que o usuário as explore, seja clicando em itens para obter mais informações sobre os mesmos, seja utilizando ferramentas comuns de visualização como *zooming*, *panning*, e até salvando a *Impressão Digital de Trânsito Urbano* em um arquivo de imagem. Utilizando essas ferramentas, o usuário pode analisar, comparar e discutir como são as rotas e como é

a acessibilidade em diferentes pontos de uma cidade, utilizando tanto o transporte público quanto o transporte privado, além de comparar as diferenças estruturais entre as rotas produzidas nas duas cidades analisadas. Resultados de algumas dessas análises, produzidas com o uso da aplicação, são apresentados no Capítulo 5. Tais resultados podem sugerir melhorias e mudanças no processo de planejamento urbano, visando solucionar problemas de mobilidade urbana.

## 4.2 Estudo de caso II: *Mapa Urbano de Acesso a Serviços Básicos*

A *Impressão Digital de Trânsito Urbano* permite a comparação da otimalidade das rotas produzidas utilizando os transportes público e privado. No entanto, para realizar a comparação dessas rotas entre si é necessário que cada rota do sistema de transporte público tenha sua correspondente (ou seja, outra rota com os mesmos pontos de origem e destino), mas produzida para o sistema de transporte privado. Com o objetivo de comparar a eficiência das rotas produzidas, e contrastar essa eficiência com outros dados da qualidade de vida de uma cidade, criamos o *Mapa Urbano de Acesso a Serviços Básicos*. A visualização é um mapa de uma cidade, dividida em regiões específicas, cuja ideia é permitir a comparação da acessibilidade a serviços básicos, como serviços de saúde e educação, utilizando os sistemas de transporte público e privado, além de relacionar tais informações com dados sobre qualidade de vida.

### 4.2.1 Integração

Para a construção do *Mapa Urbano de Acesso a Serviços Básicos* precisamos dos seguintes conjuntos de dados de uma cidade:

- Redes de circulação viária e de itinerários de ônibus.
- Dados sobre as regiões da cidade.
- Dados sobre serviços básicos de categorias específicas.
- Dados sobre a qualidade de vida em cada região da cidade.

Na criação do *Mapa Urbano de Acesso a Serviços Básicos* nos concentramos apenas no estudo e análise dos dados da cidade de Belo Horizonte. A Seção 4.1 descreve como as redes de circulação viária e de itinerários de ônibus foram construídas para



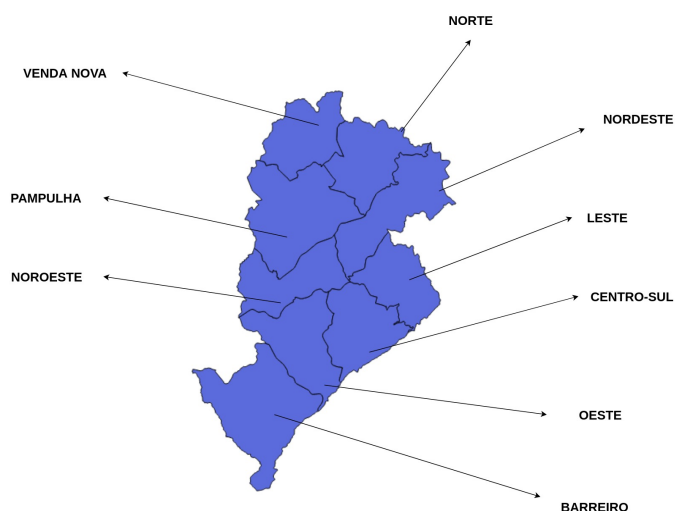


Figura 4.13: Divisão do município da cidade de Belo Horizonte nas sua regionais.

as cidades de Belo Horizonte e São Paulo, logo, esses conjuntos de dados já foram extraídos, pré-processados, integrados, padronizados e armazenados permitindo que sejam facilmente acessados e manipulados.

Uma das divisões administrativas da cidade de Belo Horizonte é a divisão em regionais. Ao todo são nove regionais: Norte, Nordeste, Leste, Barreiro, Oeste, Pampulha, Venda Nova, Noroeste e Centro-sul. Adotamos a divisão da cidade de Belo Horizonte em regionais na construção do *Mapa Urbano de Acesso a Serviços Básicos*. Além da forma de divisão da cidade, é necessário definir também as categorias de serviços a serem estudadas. No presente trabalho, nos restringimos aos dados dos serviços de saúde e educação de Belo Horizonte.

Tanto os dados das regionais, quanto dos serviços de saúde e educação de Belo Horizonte, são obtidos através do serviço WFS fornecido pela prefeitura, utilizado na Seção 4.1.1.1 para obter o polígono delimitador do município de Belo Horizonte. Utilizamos o QGIS como software cliente desse serviço, e os Shapefiles obtidos com os dados são apresentados nas Figuras 4.13, 4.14 e 4.15.

As informações sobre qualidade de vida são obtidas a partir dos dados do Censo. O Censo demográfico é um estudo estatístico referente a uma população que possibilita o recolhimento de várias informações, tais como o número de homens, mulheres, crianças e idosos, onde e como vivem as pessoas. Esse estudo é realizado normalmente de dez em dez anos, na maioria dos países. No Brasil, o responsável pelos censos demográficos é o IBGE (Instituto Brasileiro de Geografia e Estatística) que, desde 1936, quando o instituto foi criado, realiza o Censo demográfico. O último Censo realizado no Brasil foi no ano de 2010. Utilizamos os dados desse Censo na criação do *Mapa Urbano de*



Figura 4.14: Pontos referentes aos serviços de saúde da cidade de Belo Horizonte.

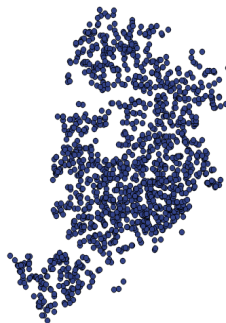


Figura 4.15: Pontos referentes aos serviços de educação da cidade de Belo Horizonte.

*Acesso a Serviços Básicos.* Dos vários índices relacionados aos municípios, o Índice de Desenvolvimento Humano Municipal (IDHM) é o que está diretamente relacionado à qualidade de vida nos ambientes urbanos. O IDHM pode ser subdividido em três dimensões:

1. **IDHM - Dimensão Educação:** É obtido através da média geométrica do subíndice de frequência escolar, com peso de  $2/3$ , e do subíndice de escolaridade, com peso de  $1/3$ .
2. **IDHM - Dimensão Longevidade:** Obtido a partir do indicador *Esperança de vida ao nascer*, através da fórmula:  $[(\text{valor observado do indicador}) - (\text{valor mínimo})] / [(\text{valor máximo}) - (\text{valor mínimo})]$ , onde os valores mínimo e máximo são 25 e 85 anos, respectivamente.
3. **IDHM - Dimensão Renda:** Obtido a partir do indicador *Renda per capita*, através da fórmula:  $[\ln(\text{valor observado do indicador}) - \ln(\text{valor mínimo})] / [\ln(\text{valor máximo}) - \ln(\text{valor mínimo})]$ , onde os valores mínimo e máximo são R\$ 8,00 e R\$ 4033,00 (a preços de agosto de 2010).

O IDHM é calculado como a média geométrica desses três indicadores, com pesos iguais. Os dados do IDHM e de suas três dimensões para a cidade de Belo Horizonte foram obtidos através do Atlas do Desenvolvimento Humano no Brasil<sup>8</sup>, que engloba o Atlas do Desenvolvimento Humano nos Municípios e o Atlas do Desenvolvimento Humano nas Regiões Metropolitanas. O Atlas é uma plataforma de consulta ao Índice de Desenvolvimento Humano Municipal (IDHM) de 5.565 municípios brasileiros, 27 Unidades da Federação (UF), 21 Regiões Metropolitanas (RM) e 3 Regiões Integradas

<sup>8</sup>[http://www.atlasbrasil.org.br/2013/pt/o\\_atlas/o\\_atlas\\_/](http://www.atlasbrasil.org.br/2013/pt/o_atlas/o_atlas_/)

de Desenvolvimento (RIDE) e suas respectivas Unidades de Desenvolvimento Humano (UDH). O Atlas traz, além do IDHM, mais de 200 indicadores de demografia, educação, renda, trabalho, habitação e vulnerabilidade, com dados extraídos dos Censos Demográficos de 1991, 2000 e 2010.

No Atlas conseguimos extrair os dados dos censos de 2000 e 2010, para todas as regionais da região metropolitana de Belo Horizonte, logo, é necessário um pré-processamento desses dados, filtrando os que nos interessam nas análises futuras. Os dados são fornecidos em um arquivo CSV, no qual realizamos a filtragem, selecionando apenas os dados das regionais do município de Belo Horizonte e referentes ao censo de 2010.

Uma tarefa importante na criação do *Mapa Urbano de Acesso a Serviços Básicos* é identificar a qual regional os dados dos serviços de saúde e educação, e do IDHM se referem. Os dados das regionais e dos serviços de saúde e educação já estão integrados, ou seja, temos uma coluna nos dados dos serviços que indica a qual regional eles pertencem. Precisamos então integrar os dados do Atlas do Desenvolvimento Humano no Brasil com os dados fornecidos pela Prefeitura de Belo Horizonte, de modo a identificar a qual regional os dados de IDHM se referem. Nos próprios dados do IDHM, existe uma coluna com o identificador da regional à qual o dado se refere. No entanto, o valor desse identificador é referente à base de dados do Atlas do Desenvolvimento Humano no Brasil. Assim, o processo de integração consiste em simplesmente alterar os valores desse campo, de modo a referenciar os identificadores das regionais de Belo Horizonte como na base de dados da Prefeitura.

Com o auxílio do QGIS, os dados dos Shapefiles com as regionais e os serviços de educação e saúde de Belo Horizonte são exportados para arquivos CSV. Para que possam ser persistidos em alguma estrutura de armazenamento de dados, todos os dados obtidos devem ser padronizados, de modo a conter apenas as colunas de dados que serão utilizadas em análises futuras. A padronização é feita da seguinte forma:

- Para as regionais: uma coluna com um identificador do tipo inteiro, uma coluna com o nome da regional, e um campo que informa a geometria, que no caso é um polígono.
- Para os serviços de saúde e educação: uma coluna com um identificador do tipo inteiro, uma coluna com o nome do serviço, uma coluna com o nome do logradouro, uma coluna com o valor do identificador da regional a qual o serviço pertence, e um campo que informa a geometria, que no caso é um ponto.

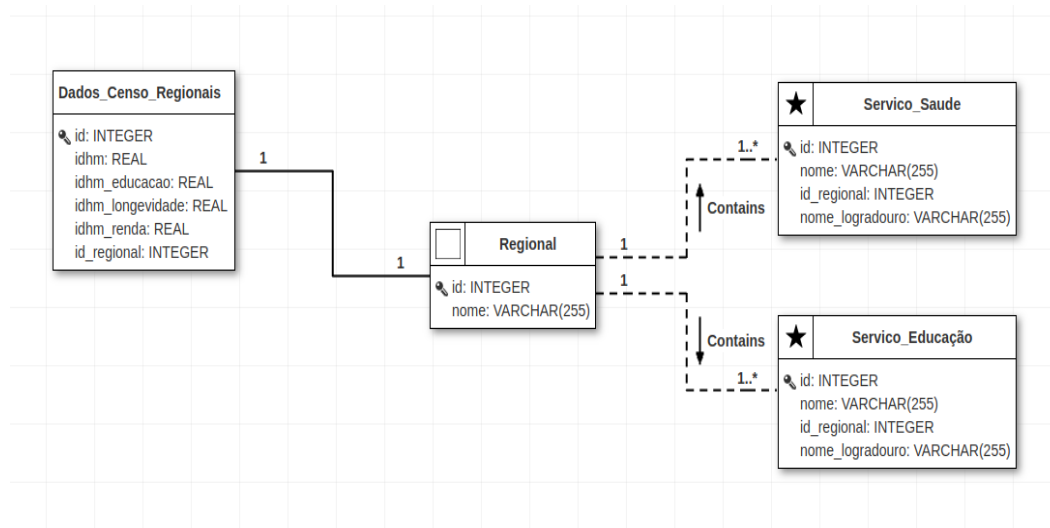


Figura 4.16: Estrutura geral e relacionamentos das tabelas utilizadas para armazenamento dos dados das regionais, dos serviços de saúde e educação, e de qualidade de vida para a cidade de Belo Horizonte.

- Para os dados de qualidade de vida: uma coluna com um identificador do tipo inteiro, uma coluna com o valor do identificador da regional a qual o dado se relaciona, e uma coluna com os valores do IDHM, e outras três colunas para os valores do IDHM nas dimensões educação, longevidade e renda.

#### 4.2.2 Armazenamento de dados

Assim como no caso da *Impressão Digital de Trânsito Urbano*, utilizamos um SGBD relacional para persistir os dados utilizados pelo *Mapa Urbano de Acesso a Serviços Básicos*, no caso, o PostGIS. Essa escolha é natural uma vez que devemos ter acesso aos dados das redes viária e de itinerário de ônibus contruídas para a *Impressão Digital de Trânsito Urbano*, e que estão armazenadas em um banco de dados PostGIS. Além disso, existe uma relação intrínseca entre os dados das regionais, e os dados de qualidade de vida e dos serviços de educação e saúde, o que sugere o uso de um SGBD relacional. Além das tabelas utilizadas para armazenamento das redes viária e de itinerários de ônibus de Belo Horizonte, cuja estrutura é apresentada na Figura 4.9, também criamos tabelas para armazenar os dados das regionais, dos serviços de saúde e educação, e os dados de qualidade de vida provenientes do censo, as estruturas dessas tabelas e as relações entre elas são apresentadas na Figura 4.16.

Como nas tabelas utilizadas para persistir os dados das redes viária e de itinerário de ônibus, os campos das tabelas que armazenam os dados das regionais, dos serviços de saúde e educação e de qualidade de vida têm relação direta com as colunas dos arquivos

CSV produzidos, o que facilita o processo de importação dos dados. Como dito, há uma relação intrínseca entre a tabela com os dados das regionais e as demais tabelas criadas, isso é evidenciado no diagrama da Figura 4.16 através dos relacionamentos de *um-para-muitos* entre a tabela com os dados das regionais e as tabelas com os dados dos serviços de saúde e educação, e de *um-para-um* entre a tabela de regionais e a tabela com os dados do censo. A persistência dos dados em tabelas do PostGIS facilita sua organização, acesso e consulta pelos algoritmos criados na construção do *Mapa Urbano de Acesso a Serviços Básicos*.

### 4.2.3 Prestação de serviços

Assim como para a *Impressão Digital de Trânsito Urbano*, a construção do *Mapa Urbano de Acesso a Serviços Básicos* se baseia na criação de rotas utilizando as redes viária e de itinerários de ônibus de uma cidade. Como a ideia é comparar o acesso aos serviços de uma categoria específica, de cada regional, utilizando os sistemas de transporte público e privado, definimos que a origem das rotas criadas é o centroide do polígono representativo de cada regional, e os destinos são os pontos que representam os serviços. Logo, uma forma simples de gerar tais rotas é, para cada serviço:

- **Rota no sistema de transporte privado:** Encontrar os pontos de interseção de via mais próximo do centroide da regional, e do serviço de destino, e calcular a rota como o caminho mínimo entre tais pontos na rede viária.
- **Rota no sistema de transporte público:** Encontrar os pontos de parada de ônibus mais próximo do centroide da regional, e do serviço de destino, e calcular a rota como o caminho mínimo entre tais pontos na rede de itinerários de ônibus.

No entanto, as rotas criadas dessa forma não podem ser comparadas corretamente, uma vez que, elas não possuem necessariamente os mesmos pontos de origem e destino. Para que isso ocorra, devemos fracionar as vias inicial e final das rotas produzidas para o sistema de transporte privado de modo que elas se iniciem exatamente no seu ponto de origem, e terminem exatamente no seu ponto de destino, e incluir rotas de baldeação a pé nas rotas produzidas para o sistema de transporte público. Essas rotas de baldeação devem ligar: o centroide ao ponto de ônibus mais próximo dele, e o ponto de ônibus mais próximo do serviço ao serviço de destino.

Como dito, o algoritmo de fracionamento das rotas produzidas utilizando a rede viária de uma cidade consiste em subdividir as vias inicial e final dessas rotas. Isso é feito da seguinte forma:

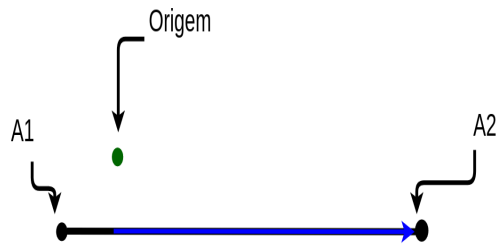


Figura 4.17: Projeção do ponto de origem posicionada entre o ponto de origem (A1) e destino (A2) da via.



Figura 4.18: Projeção do ponto de origem posicionada no ponto de origem (A1) da via.

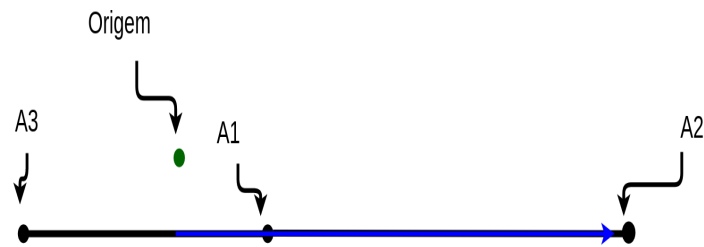


Figura 4.19: Projeção do ponto de origem posicionada fora da via, antes do ponto de origem (A1) da via.

1. Considerando que a via inicial de uma rota liga os pontos A1 e A2, então o fracionamento dessa via depende do posicionamento do ponto de origem da rota. Existem três posicionamentos que interessam:

- A projeção do ponto de origem da rota está localizada entre A1 e A2, isso é mostrado na Figura 4.17. Nesse caso, a via é fracionado normalmente, partindo da fração ocupada pela projeção do ponto de origem na via, e indo até o ponto final da via (A2). O resultado desse fracionamento é indicado pela seta azul na Figura 4.17.
- A projeção do ponto de origem da rota está localizada em A1, como é apresentado na Figura 4.18. Nesse caso, a via não é fracionada, e é mantida em sua forma original na rota, como indicado pela seta azul presente na Figura 4.18.
- A projeção do ponto de origem da rota está localizada em uma reta em que A1 é o ponto de destino, ou seja, a projeção está geolocalizada antes de A1, como apresentado na Figura 4.19. Nesse caso, devemos primeiramente identificar a via na qual a projeção está localizada, ela obrigatoriamente é a via mais próxima do ponto de origem e que tem o ponto A1 como destino. Uma vez identificada, tal via é fracionada, criando uma subvia que vai da

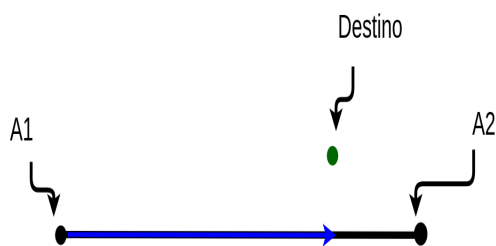


Figura 4.20: Projeção do ponto de destino posicionada entre o ponto de origem (A1) e destino (A2) da via.

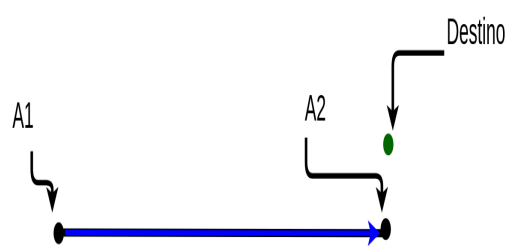


Figura 4.21: Projeção do ponto de destino posicionada no ponto de destino (A2) da via.

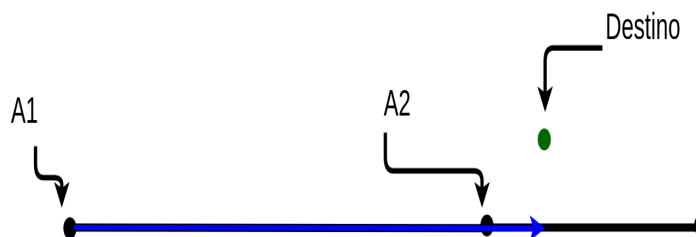


Figura 4.22: Projeção do ponto de destino posicionada fora da via, após o ponto de destino (A2) da via.

projeção da origem na via até A1. Essa subvia e a via que liga A1 a A2 são adicionadas à rota, como apresentado na Figura 4.19.

2. Sendo a via que liga os pontos A1 e A2 a via final de uma rota, então o fracionamento dessa via depende do posicionamento do ponto de destino da rota. Existem três posicionamentos devem ser considerados:

- A projeção do ponto de destino da rota está localizada entre A1 e A2, isso é mostrado na Figura 4.20. Nesse caso, da mesma forma que para o ponto de origem, a via é fracionado normalmente, partindo da fração ocupada pela projeção do ponto de destino na via, e indo até o ponto final da via (A2).
- A projeção do ponto de destino da rota está localizada em A2, como é apresentado na Figura 4.21. Nesse caso, a via não é fracionada, e é utilizada sem alterações na rota, como indicado na Figura 4.21.
- A projeção do ponto de destino da rota está localizada em uma reta em que A2 é o ponto de origem, ou seja, a projeção está geolocalizada após A2, como apresentado na Figura 4.22. Nesse caso, de modo análogo à origem, devemos primeiramente identificar a via na qual a projeção está localizada,

ela é a via mais próxima do ponto de destino e que tem o ponto A2 como origem. Essa via então é fracionada, criando uma subvia que vai de A2 até a projeção do destino na via. Essa subvia e a via que liga A1 a A2 são adicionadas à rota, como é mostrado na Figura 4.22.

Posicionamentos em que o ponto de origem está localizado no ponto de destino A2, ou após o ponto de destino A2, e o ponto de destino está localizado em A1, ou antes de A1, não são considerados por não fazerem sentido. Em todos esses casos, a via que liga A1 a A2 não seria utilizada na rota, e assim, as nossas suposições iniciais seriam inválidas.

Utilizamos o método descrito anteriormente para corrigir as rotas produzidas utilizando o sistema de transporte privado. Para as rotas criadas utilizando o sistema de transporte público, necessitamos de rotas de baldeação a pé. Como dito, devemos criar duas dessas rotas: uma ligando o *centroid* ao ponto de ônibus mais próximo dele, e outra ligando o ponto de ônibus mais próximo do serviço ao serviço de destino. Em ambos os casos, as rotas são criadas da mesma forma:

1. Calculamos a rota inicial como o caminho mínimo entre os nós da rede viária mais próximos do ponto de origem (centroide e ponto de ônibus mais próximo do serviço de destino) e de destino (ponto de ônibus mais próximo do centroide e serviço de destino).
2. Corrigimos essa rota inicial utilizando o método descrito, de modo que ela inicie exatamente no seu ponto de origem e termine exatamente no seu ponto de destino.
3. Dado o tamanho em quilômetros dessa rota, calculamos o tempo que uma pessoa a pé leva para percorrê-la, considerando que a velocidade média de uma pessoa caminhando é 5 quilômetros por hora.

As rotas de baldeação a pé são então adicionadas ao trajeto calculado utilizando os itinerários de ônibus. Após essas correções, finalmente temos rotas para cada um dos serviços de destino, que iniciam e terminam nos mesmos pontos, e que utilizam os dois sistemas de transporte analisados. Assim, para comparar tais rotas, criamos o *Indicador de Ineficiência do Transporte Público*. Sendo *len* uma função que calcula o tamanho de uma rota em metros, e *time* outra função que determina o tempo em segundos para percorrer uma rota, o *Indicador de Ineficiência do Transporte Público* é calculado como:



$$\begin{aligned}
 \mathbf{di} &= \left( \frac{\text{len}(\text{Rota}_{\text{Transporte\_Publico}})}{\text{len}(\text{Rota}_{\text{Transporte\_Privado}})} \right) \\
 \mathbf{ti} &= \left( \frac{\text{time}(\text{Rota}_{\text{Transporte\_Publico}})}{\text{time}(\text{Rota}_{\text{Transporte\_Privado}})} \right) \\
 \mathbf{indicador} &= \mathbf{di} * \mathbf{ti}
 \end{aligned}$$

A partir das fórmulas acima, podemos concluir que:

1. Caso  $\mathbf{di} > 1$ , a rota criada utilizando o transporte público é mais eficiente que a gerada pelo transporte privado com relação à distância percorrida. O mesmo vale caso  $\mathbf{ti} > 1$ , porém, analisando o tempo para percorrer tais rotas.
2. Caso  $0 \leq \mathbf{di} < 1$ , a rota criada utilizando o transporte público é menos eficiente que a gerada pelo transporte privado com relação à distância percorrida. O mesmo vale caso  $0 \leq \mathbf{ti} < 1$ , porém, analisando a variável tempo.
3. O **indicador** é calculado como o produto das duas frações, visando contrabalançar casos em que um sistema de transporte é mais eficiente em um quesito, mas pior em outro. Como exemplo, temos os corredores de ônibus, nos quais, apesar de os ônibus percorrerem maiores distâncias em seus itinerários, eles muito provavelmente os percorre em menor tempo, por terem faixas exclusivas.

O indicador é calculado para cada serviço de uma regional, uma vez que, para cada serviço, existe um par de rotas que utilizam os dois sistemas de transporte. Além de calcular o indicador para cada serviço de destino, calculamos também a mediana desse indicador para todos os serviços de uma regional, o tempo médio de viagem e a distância média percorrida em cada um dos sistemas de transporte para cada regional.

Todos os algoritmos e indicadores discutidos são responsáveis pela análise e processamento dos dados. Como foi feito na *Impressão Digital de Trânsito Urbano*, os algoritmos e indicadores do *Mapa Urbano de Acesso a Serviços Básicos* foram criados como *webservices*, cujos resultados são consumidos pelo serviço final, que interfaceia diretamente com o usuário. Na interface disponibilizada por esse serviço, o usuário deve selecionar o tipo de serviço básico (educação ou saúde) que será utilizado na consulta, e após clicar no botão para gerar o mapa de mobilidade, é apresentado o mapa coroplético da cidade de Belo Horizonte apresentado na Figura 4.23, nesse mapa, as cores codificam o valor do IDH de cada regional, de acordo com a legenda apresentada. A visualização é interativa, assim, ao clicar sobre uma região são apresentados:

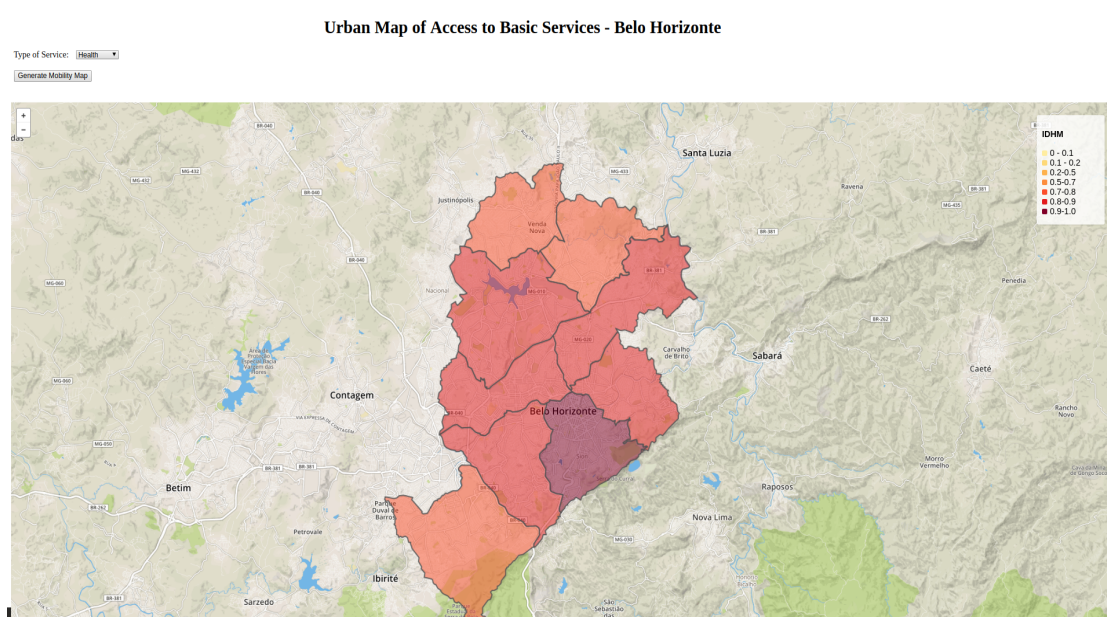


Figura 4.23: Mapa coroplético da cidade de Belo Horizonte, as cores de cada regional indicam o valor do IDH da regional.

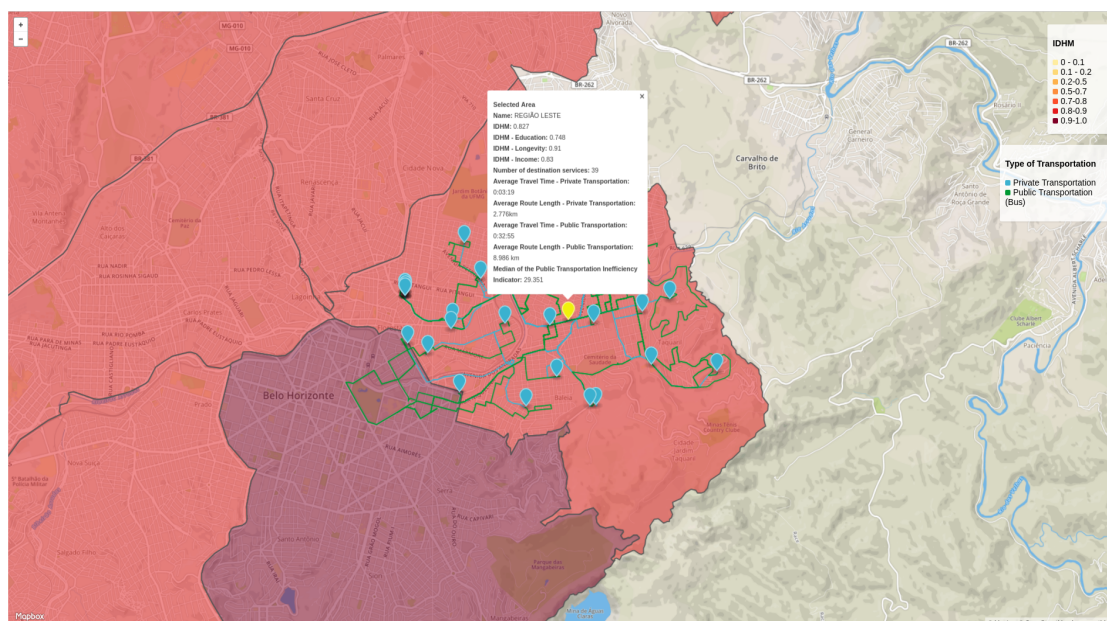


Figura 4.24: Resultados apresentados pelo sistema para a Regional Leste de Belo Horizonte.

- As rotas utilizando os sistemas de transporte público e privado para cada um dos serviços da categoria selecionada.
- Para a regional são mostrados: Nome, valores do IDH e de suas dimensões (educação, longevidade e renda), número de serviços na regional, o tempo médio de

viagem utilizando o transporte público e o privado, a distância média em quilômetros percorrida nas rotas dos transportes público e privado, e o valor da mediana do *Indicador de Ineficiência do Transporte Público* das rotas calculadas.

- Para cada serviço de destino da categoria selecionada temos: Seu nome, endereço, tempo de viagem usando o transporte público e o privado, distância percorrida em quilômetros pelas rotas, e o valor do *Indicador de Ineficiência do Transporte Público* para o par de rotas relacionadas.

Esses resultados estão presentes na Figura 4.24. Da mesma forma que a *Impressão Digital de Trânsito Urbano*, o *Mapa Urbano de Acesso a Serviços Básicos* permite a análise, comparação e discussão de como são as rotas e como é a acessibilidade a diferentes serviços de uma cidade. Ao contrastar as rotas e indicadores produzidos com os dados do IDHM, a visualização permite ao planejador urbano identificar correlações e tentar corrigir possíveis problemas de acesso e mobilidade.

---

**Algoritmo 6** Algoritmo para construção das arestas da rede de itinerários de ônibus de uma cidade a partir dos dados pré-processados dos arquivos *routes.txt*, *trips.txt*, *stop\_times.txt*, *stops.txt* e *shapes.txt*

---

**Entrada:**

Rotas, Viagens, Shapes, Paradas, Trajetos: dicionários com os dados pré-processados dos arquivos *routes.txt*, *trips.txt*, *shapes.txt*, *stops.txt* e *stop\_times.txt*.

**Saída:**

Itinerarios : Lista de tuplas que armazenam os dados de cada aresta da rede de itinerários de ônibus.

```

1: procedimento CONSTROI ARESTAS ITINERARIOS
2:   Itinerarios  $\leftarrow \emptyset$ 
3:   para cada chave  $\in$  chaves do dicionário Trajetos faça
4:     id_rota  $\leftarrow$  Viagens[chave][0]
5:     id_rota  $\leftarrow$  Viagens[chave].id_rota
6:     id_shape  $\leftarrow$  Viagens[chave].id_shape
7:     nome_rota  $\leftarrow$  Rotas[id_rota].nome
8:     trajeto  $\leftarrow$  Trajetos[chave]

9:     para i em 0..len(trajeto) - 1 faça
10:      tempo_trecho_trajeto  $\leftarrow$  trajeto[i + 1].horario_chegada -
                                trajeto[i].horario_saida
11:      id_parada_origem  $\leftarrow$  trajeto[i].id_parada
12:      id_parada_destino  $\leftarrow$  trajeto[i + 1].id_parada
13:      pontos_no_shape  $\leftarrow$  Shapes[id_shape]

14:      ponto_shape_proximo_origem  $\leftarrow$  obtempontoMaisProximoDoShape(
                                pontos_no_shape,
                                Paradas[id_parada_origem])
15:      ponto_shape_proximo_destino  $\leftarrow$  obtempontoMaisProximoDoShape(
                                pontos_no_shape,
                                Paradas[id_parada_destino])
16:      linha_trecho_itinerario  $\leftarrow$  criaLinhaLigandoConjuntoPontos(
                                pontos_no_shape,
                                ponto_shape_proximo_origem,
                                ponto_shape_proximo_destino)
17:      Itinerarios += (nome_rota, id_parada_origem, id_parada_destino,
                       tempo_trecho_trajeto, linha_trecho_itinerario)

18:     fim para
19:   fim para
20:   retorna Itinerarios
21: fim procedimento

```

---

---

**Algoritmo 7** Algoritmo para fazer a associação entre as vias da rede viária e os pontos de parada de ônibus

---

**Entrada:**

**Paradas:** dicionário com os dados dos pontos de paradas de ônibus contidos no arquivo *stops.txt* pré-processados.

**Vias\_Rede\_Viaria:** dicionário com os dados das vias da rede viária da cidade.

**Saída:**

**Vias\_Pontos\_Parada :** Dicionário do tipo chave-valor. Armazena para cada via da rede viária uma lista de tuplas que informam o identificador do ponto de parada e a fração que a projeção desse ponto divide a via.

```
1: procedimento CRIA ASSOCIACAO VIAS REDE VIARIA PONTOS DE PARADA
2:   Vias_Pontos_Parada  $\leftarrow$  {}
3:   para cada chave  $\in$  chaves do dicionário Paradas faça
4:     id_ponto_parada  $\leftarrow$  chave
5:     via_proxima_ponto_parada  $\leftarrow$  obtemArestaProximaPontoParada(
6:       Vias_Rede_Viaria,
7:       id_ponto_parada)
8:     id_via  $\leftarrow$  via_proxima_ponto_parada.id
9:     fracao_ponto_via  $\leftarrow$  via_proxima_ponto_parada.fracao
10:    Vias_Pontos_Parada[id_via] += (id_ponto_parada,
11:      fracao_ponto_via)
12:   fim para
13:   retorna Vias_Pontos_Parada
14: fim procedimento
```

---

---

**Algoritmo 8** Algoritmo para a construção do grafo correspondente à rede de circulação híbrida, resultado da integração da rede viária com um *feed* GTFS

---

**Entrada:**

- Vias\_Redde\_Viaria: dicionário com os dados das vias da rede viária da cidade.  
 Vias\_Pontos\_Parada: Dicionário com as associações entre vias da rede viária e a lista de pontos de parada localizadas nas mesmas.  
 i: índice do identificador do nó origem das vias nas tuplas do dicionário Vias\_Redde\_Viaria.  
 k: índice do identificador do nó destino das vias nas tuplas do dicionário Vias\_Redde\_Viaria.  
 l: índice do identificador do das vias nas tuplas do dicionário Vias\_Redde\_Viaria.

**Saída:**

Rede\_Hibrida : Grafo com conjunto de vértices  $V$  e de arestas  $E$ , que representa a rede de circulação híbrida.

```

1: procedimento CRIA REDE DE CIRCULACAO HIBRIDA
2:   Rede_Hibrida  $\leftarrow (V, E)$ 
3:   para cada via  $\in$  Vias_Redde_Viaria faça
4:     id_via  $\leftarrow$  via[l]
5:     origem  $\leftarrow$  via[j]
6:     destino  $\leftarrow$  via[k]
7:     paradas_na_via  $\leftarrow$  Vias_Pontos_Parada[id_via]

8:     se paradas_na_via  $\neq \emptyset$  então
9:       paradas_na_via.enfileira((origem, 0.0))
10:      paradas_na_via.empilha((destino, 1.0))
11:      paradas_na_via.ordenaEntradasPeloValorFracao()

12:      para x em 0..len(paradas_na_via) - 1 faça
13:        parada_atual  $\leftarrow$  paradas_na_via[x]
14:        parada_seguinte  $\leftarrow$  paradas_na_via[x + 1]
15:        subvia  $\leftarrow$  subdivideVia(
16:          via,
17:          parada_atual.fracao,
18:          parada_seguinte.fracao)
19:         $V_{\text{Rede\_Hibrida}} \leftarrow V_{\text{Rede\_Hibrida}} \cup \{\text{parada\_atual}\}$ 
20:         $V_{\text{Rede\_Hibrida}} \leftarrow V_{\text{Rede\_Hibrida}} \cup \{\text{parada\_seguinte}\}$ 
21:         $E_{\text{Rede\_Hibrida}} \leftarrow E_{\text{Rede\_Hibrida}} \cup \{\text{subvia}\}$ 
22:      fim para
23:    senão
24:       $V_{\text{Rede\_Hibrida}} \leftarrow V_{\text{Rede\_Hibrida}} \cup \{\text{origem}\}$ 
25:       $V_{\text{Rede\_Hibrida}} \leftarrow V_{\text{Rede\_Hibrida}} \cup \{\text{destino}\}$ 
26:       $E_{\text{Rede\_Hibrida}} \leftarrow E_{\text{Rede\_Hibrida}} \cup \{\text{via}\}$ 
27:    fim se
28:  fim para
29:  retorna Rede_Hibrida
30: fim procedimento

```

---

---

**Algoritmo 9** Algoritmo para construção das arestas da rede de itinerários de ônibus a partir dos dados pré-processados dos arquivos *routes.txt*, *trips.txt*, *stop\_times.txt*, *stops.txt* e da rede de circulação híbrida criada com o Algoritmo 8

---

**Entrada:**

Rotas, Viagens, Paradas, Trajetos: dicionários com os dados pré-processados dos arquivos *routes.txt*, *trips.txt*, *stops.txt* e *stop\_times.txt*.

Rede\_Hibrida: Grafo que representa a rede de circulação híbrida.

**Saída:**

Itinerarios: Lista de tuplas que armazenam os dados de cada aresta da rede de itinerários de ônibus.

```

1: procedimento CONSTROI ARESTAS ITINERARIOS USANDO REDE HIBRIDA
2:   Itinerarios  $\leftarrow \emptyset$ 
3:   para cada chave  $\in$  chaves do dicionário Trajetos faça
4:     id_rota  $\leftarrow$  Viagens[chave].id_rota
5:     id_shape  $\leftarrow$  Viagens[chave].id_shape
6:     nome_rota  $\leftarrow$  Rotas[id_rota].nome
7:     trajeto  $\leftarrow$  Trajetos[chave]

8:     para i em 0..len(trajeto) - 1 faça
9:       tempo_trecho_trajeto  $\leftarrow$  trajeto[i + 1].horario_chegada -
                                trajeto[i].horario_saida
10:      id_parada_origem  $\leftarrow$  trajeto[i].id_parada
11:      id_parada_destino  $\leftarrow$  trajeto[i + 1].id_parada

12:      linha_trecho_itinerario  $\leftarrow$  DIJKSTRA(
                                Rede_Hibrida,
                                Paradas[id_parada_origem],
                                Paradas[id_parada_destino])

13:      se linha_trecho_itinerario == NIL então
14:        linha_trecho_itinerario  $\leftarrow$  criaLinhaEntreDoisPontos(
                                Paradas[id_parada_origem],
                                Paradas[id_parada_destino])

15:      fim se
16:      Itinerarios += (nome_rota, id_parada_origem, id_parada_destino,
                       tempo_trecho_trajeto, linha_trecho_itinerario)

17:   fim para
18: fim para
19:   retorna Itinerarios
20: fim procedimento

```

---





# Capítulo 5

## Resultados

Neste Capítulo são apresentados resultados e análises feitas utilizando as duas aplicações finais contruídas nos casos de estudo apresentados, ou seja, a *Impressão Digital de Trânsito Urbano* e o *Mapa Urbano de Acesso a Serviços Básicos*. Os resultados corroboram a validade dos serviços criados no sentido de auxiliar o planejador urbano nas solução de problemas de acesso e mobilidade urbana.

### 5.1 Estudo de caso I: *Impressão Digital de Trânsito Urbano*

A *Impressão Digital de Trânsito Urbano* permite ao usuário comparar a acessibilidade utilizando os sistemas de transporte público e privado a partir de um ponto qualquer de uma cidade. Assim, para cada uma das cidades disponíveis selecionamos dois pontos de interesse, um na área central da cidade, e outro em uma área periférica. Dessa forma, podemos comparar tanto qualitativamente (através das visualizações produzidas), quanto quantitativamente (através do *Indicador de Ineficiência da Rota*) como é o acesso em diferentes pontos da cidade, além de permitir a comparação entre as cidades analisadas. Em todas as análises feitas, buscamos pelos pontos que são alcançáveis a partir do ponto selecionado no intervalo de 10 a 15 minutos.

Para a cidade de Belo Horizonte, o ponto de interesse selecionado para representar uma área central da cidade é a Praça Sete de Setembro, localizada no centro da capital mineira, é a praça mais movimentada da cidade e marco zero do seu hipercentro. As Figuras 5.1 e 5.2 apresentam as visualizações produzidas pela *Impressão Digital de Trânsito Urbano*, utilizando a Praça Sete de Setembro como origem das rotas, para os sistemas de transporte privado e público respectivamente.

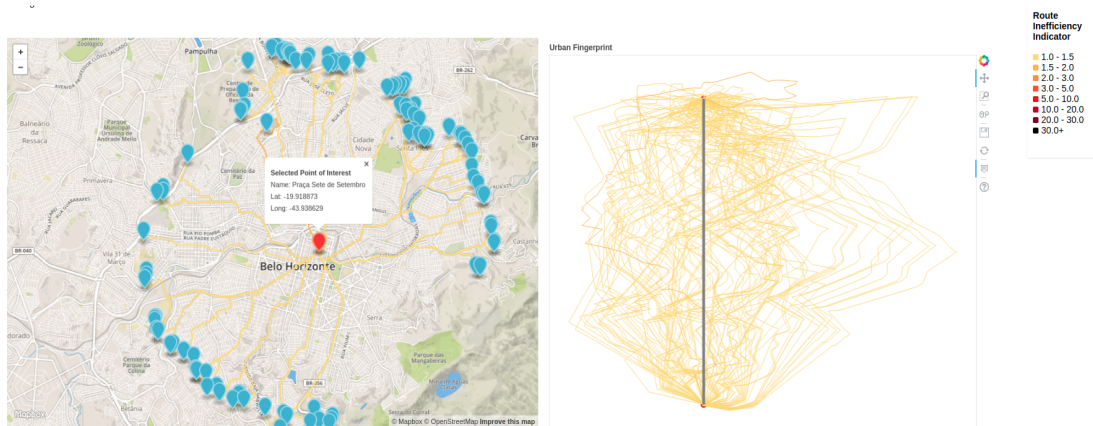


Figura 5.1: Resultado final da *Impressão Digital de Trânsito Urbano* tendo a Praça Sete de Setembro como origem das rotas. Rotas produzidas utilizando o sistema de transporte privado.

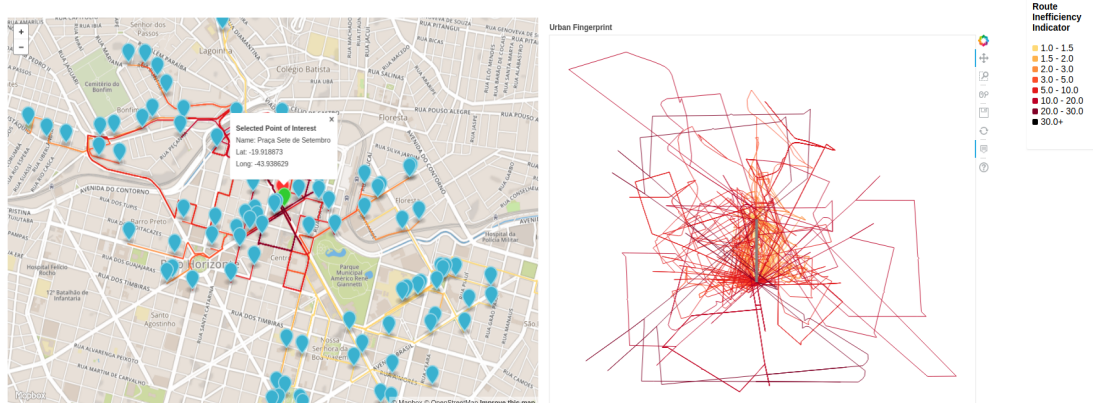


Figura 5.2: Resultado final da *Impressão Digital de Trânsito Urbano* tendo a Praça Sete de Setembro como origem das rotas. Rotas criadas a partir dos dados do sistema de transporte público.

O ponto de interesse representativo de uma área periférica da cidade de Belo Horizonte é a unidade da Pontifícia Universidade Católica de Minas Gerais (PUC - MG) no bairro Coração Eucarístico. A Figura 5.3 apresenta os resultados produzidos pela *Impressão Digital de Trânsito Urbano* utilizando os dados da rede viária de Belo Horizonte para o cálculo das rotas. Já a Figura 5.4 apresenta o resultado final utilizando os dados da rede de itinerários de ônibus de Belo Horizonte.

Para a cidade de São Paulo, o ponto central selecionado foi a Catedral Municipal de São Paulo, localizada na Praça da Sé, na zona central da cidade. O ponto periférico escolhido foi a Arena Corinthians, estádio de futebol localizado no distrito de Itaquera, zona leste da cidade. As Figuras 5.5 e 5.6 apresentam as visualizações criadas tendo como origem a Catedral Municipal de São Paulo, e utilizando as rotas calculadas com

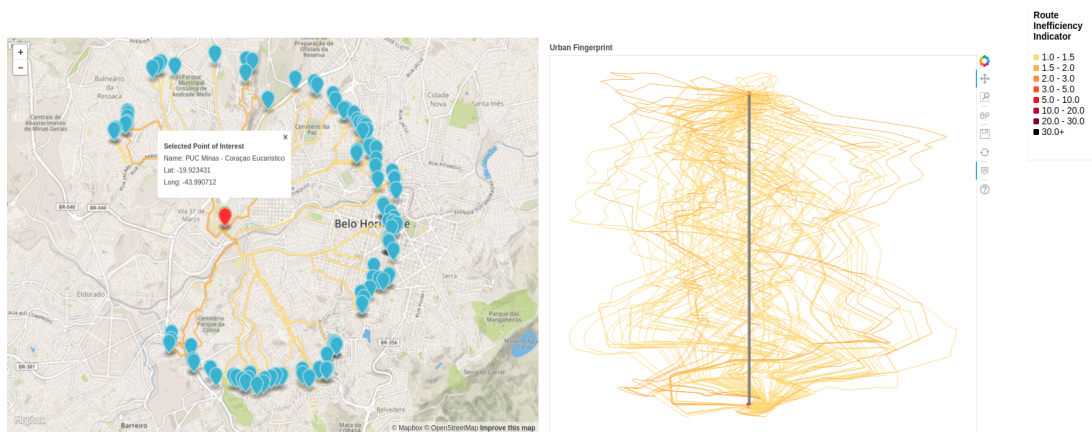


Figura 5.3: Visualizações produzidas tendo a unidade da PUC-MG no bairro Coração Eucarístico como origem. Rotas produzidas utilizando o sistema de transporte privado.

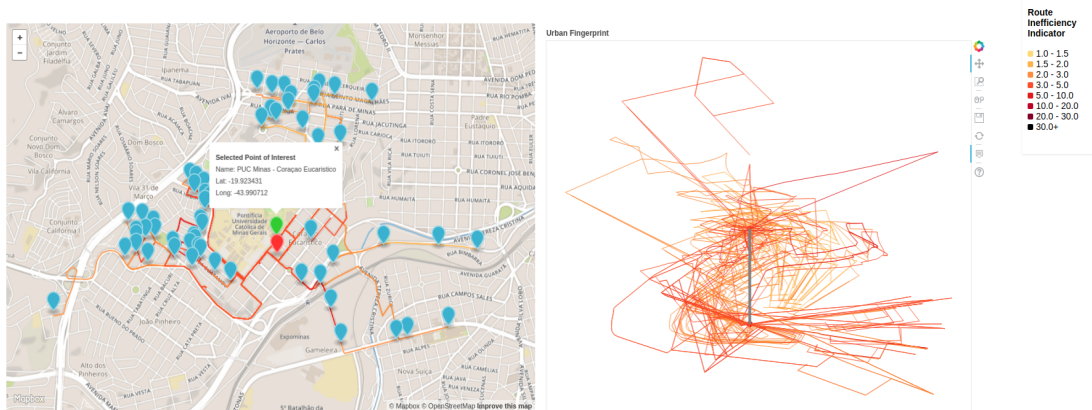


Figura 5.4: Visualizações produzidas tendo a unidade da PUC-MG no bairro Coração Eucarístico como origem. As rotas foram criadas a partir dos dados o sistema de transporte público.

os dados do sistema de transporte privado e público respectivamente. As Figuras 5.7 e 5.8 apresentam os mesmos resultados, porém tendo como origem a Arena Corinthians.

A Tabela 5.1 apresenta os valores das medianas do *Indicador de Ineficiência da Rota* para as rotas produzidas em cada um dos cenários anteriormente citados. Analisando as visualizações criadas e os valores das médias para o *Indicador de Ineficiência da Rota* observamos o seguinte:

- Utilizando o sistema de transporte privado, é possível alcançar pontos mais distantes no mesmo intervalo de tempo, e as rotas criadas são mais próximas da ótima. Isso pode ser percebido tanto nas visualizações apresentadas, quanto nos valores das médias do *Indicador de Ineficiência da Rota* presentes na Tabela 5.1. Para o caso das visualizações, podemos claramente perceber o maior alcance das

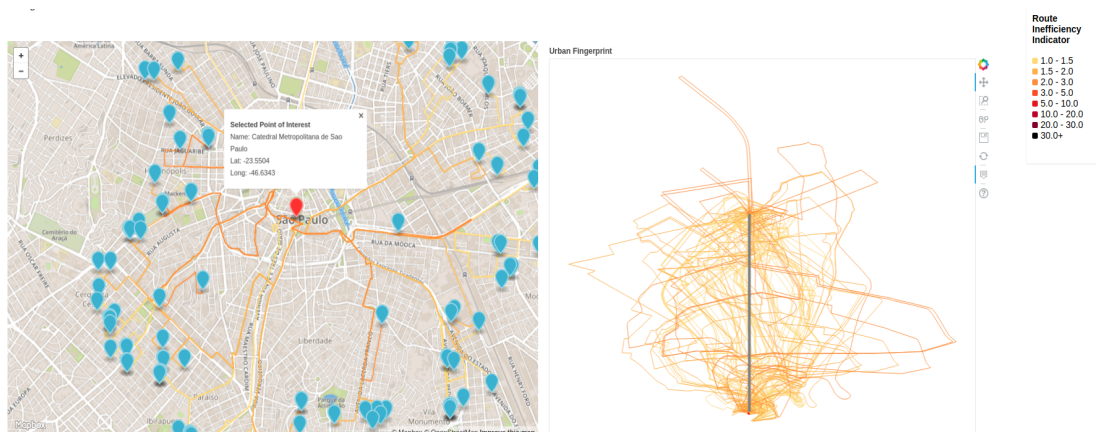


Figura 5.5: Visualizações da *Impressão Digital de Trânsito Urbano* tendo a Catedral Municipal de São Paulo como origem. Rotas produzidas com os dados do sistema de transporte privado.

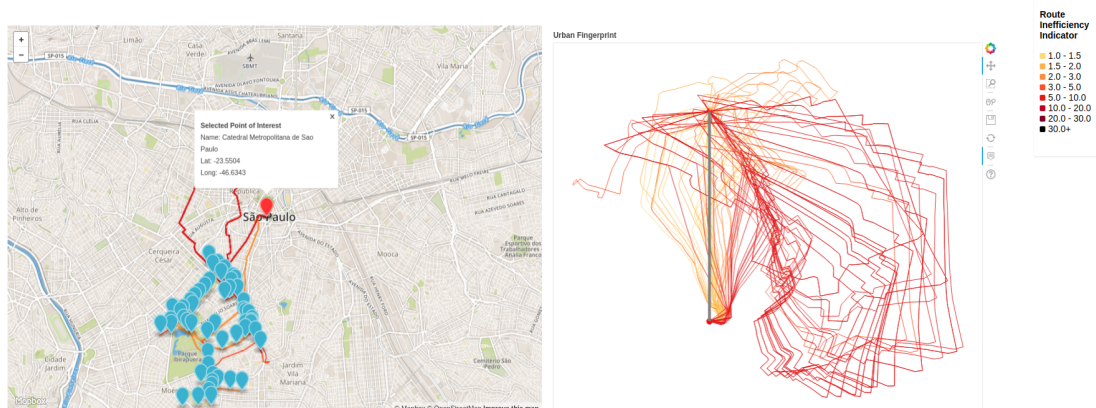


Figura 5.6: Visualizações da *Impressão Digital de Trânsito Urbano* tendo a Catedral Municipal de São Paulo como origem. Rotas produzidas utilizando os dados do sistema de transporte público.

rotas produzidas pelo sistema de transporte privado, em todos os casos estudados. Além disso, as rotas do sistema de transporte privado evidenciam sua proximidade com a otimalidade tanto na forma (são menos curvas e mais próximas de uma linha reta), quanto na cor atribuída à elas (que corresponde ao valor do *Indicador de Ineficiência da Rota*). Interessante notar também que por serem mais distantes da rota ótima, as rotas criadas com o sistema de transporte público são em geral deformadas na criação da *Impressão Digital de Trânsito Urbano*, isso ocorre no processo de redimensionamento das rotas. Como essas rotas tem formas muito diferentes, muitas vezes, a escolha da rota de referência cria fatores de redimensionamento que aumentam ou reduzem muito certas rotas, o que gera as distorções percebidas.

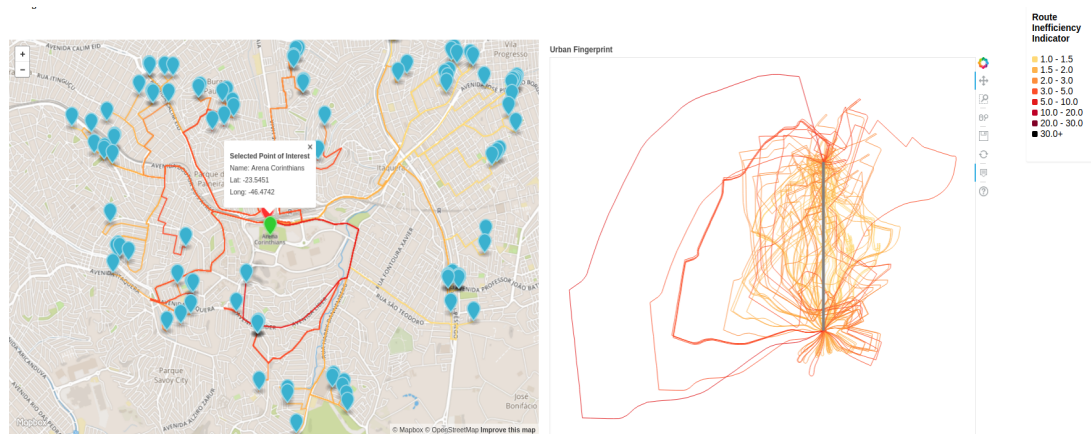


Figura 5.7: Resultado final da *Impressão Digital de Trânsito Urbano* com a Arena Corinthians como origem. Rotas produzidas utilizando o sistema de transporte privado.

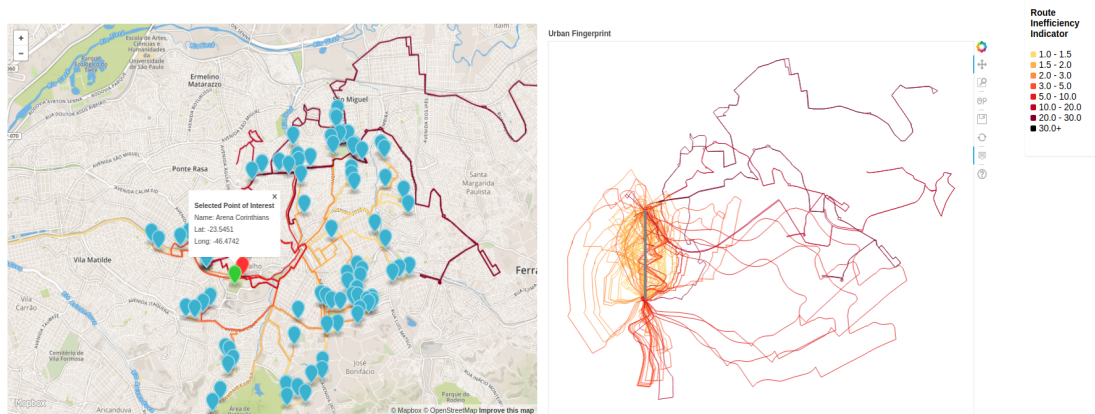


Figura 5.8: Resultado final da *Impressão Digital de Trânsito Urbano* com a Arena Corinthians como origem. Rotas produzidas a partir dos dados do sistema de transporte público.

- As rotas criadas para pontos centrais das cidades analisadas, para ambos os sistemas de transporte são, em geral, mais próximas da ótima do que as criadas para pontos periféricos. Isso é evidenciado pelos dados da Tabela 5.1, e pode ser percebido pela predominância de rotas com cores mais avermelhadas (com *Indicador de Ineficiência da Rota* mais alto) sobretudo nas rotas do sistema de transporte público, nas visualizações produzidas para pontos periféricos das cidades.
- Comparativamente, as rotas produzidas em ambos os sistemas de transporte, tanto para os pontos de interesse centrais, quanto para os periféricos para a cidade de São Paulo são mais próximas da ótima que as produzidas para a cidade de Belo Horizonte. Isso pode ser o resultado de vários fatores, um dos quais é o

Tabela 5.1: Valores das medianas do *Indicador de Ineficiência da Rota* para os sistemas de transporte público e privado, em cada um dos pontos de interesse selecionados.

Ponto de Interesse	Mediana <i>Indicador de Ineficiência da Rota - Transporte Privado</i>	Mediana <i>Indicador de Ineficiência da Rota - Transporte Público</i>
Praça Sete de Setembro - BH	1.20	2.23
Catedral Municipal de São Paulo - SP	1.42	1.69
PUC - Coração Eucarístico - BH	1.27	2.25
Arena Corinthians - SP	1.71	2.02

fato da cidade de São Paulo ter um sistema de metrô muito mais amplo, com 5 linhas, 80 estações e 90 km de trilhos, servindo uma parte considerável da cidade, e ajudando a evitar o uso dos ônibus. Por outro lado, a cidade de Belo Horizonte possui um sistema metroviário bem menor, com apenas uma linha, 19 estações e 28 km de extensão, fazendo dos ônibus os principais meios de transporte público. A geometria do sistema de ônibus leva a rotas muito longas e complicadas, para permitir que as vizinhanças periféricas sejam cobertas.

Os resultados e análises apresentados sugerem mudanças estruturais sobretudo no sistema de transporte público das cidades estudadas. O fato de as rotas produzidas pelo sistema privado estarem muito mais próximas do ideal já era esperado, porém a ineficiência do sistema de transporte público é evidente, já que o *Indicador de Ineficiência da Rota* mostra que o cidadão leva até 123% mais tempo para chegar a um destino usando o sistema público de transporte. Considerando que os ônibus operam com cerca de metade da velocidade do tráfego regular, essa ineficiência de distância pode resultar em viagens que duram até quatro vezes mais se comparadas com o uso de um veículo particular. Não é de admirar que o número de passageiros no transporte público de massa esteja diminuindo. Carvalho & Pereira [6] apontam queda de cerca de 30% na demanda de transporte urbano utilizando ônibus nas cidades brasileiras no período de 1995 a 2013. Gomide et al. [13] indicam a perda de 18,1% na demanda de passageiros entre 2013 e 2016. Essa redução no uso de transporte público e aumento no uso de transporte individual apresenta consequências desastrosas em termos de congestionamento e poluição do ar. Assim, para ambas as cidades, são necessárias correções nos itinerários para suprir essas deficiências. Além disso, melhorar a acessibilidade nas regiões periféricas é uma medida necessária para a melhoria da mobilidade urbana.

Tabela 5.2: Tabela comparativa entre os dados de qualidade de vida (IDH) e o valor da mediana do *Indicador de Ineficiência do Transporte Público* por regional para os serviços de saúde de Belo Horizonte.

Regional	IDH	IDH - Educação	IDH - Longevidade	IDH - Renda	Mediana - <i>Indicador de Ineficiência do Transporte Público</i>
Venda Nova	0.755	0.683	0.860	0.732	27.39
Norte	0.754	0.679	0.862	0.733	19.5
Leste	0.827	0.748	0.910	0.830	16.52
Barreiro	0.744	0.672	0.852	0.720	14.32
Nordeste	0.801	0.719	0.895	0.800	13.00
Noroeste	0.818	0.744	0.907	0.811	10.59
Pampulha	0.853	0.793	0.911	0.859	10.42
Oeste	0.839	0.757	0.910	0.857	9.29
Centro-sul	0.914	0.823	0.933	0.995	8.15

## 5.2 Estudo de caso II: *Mapa Urbano de Acesso a Serviços Básicos*

O principal objetivo do *Mapa Urbano de Acesso a Serviços Básicos* é permitir a comparação da eficiência das rotas produzidas pelos sistemas de transporte público e privado, revelando assim problemas na acessibilidade a serviços básicos para a população, e contrastar tais dados com informações sobre qualidade de vida. No caso de estudo criado, nos concentramos no uso dos dados da cidade de Belo Horizonte, e no acesso a serviços de saúde (hospitais, postos de saúde, clínicas, farmácias, etc) e de educação (escolas, faculdades, creches, etc). Com o objetivo de quantificar a diferença entre as rotas produzidas, criamos o *Indicador de Ineficiência do Transporte Público*, que possui esse nome justamente porque ele indica o quanto o sistema de transporte público é mais eficiente ou menos eficiente que o transporte privado, considerando tanto o tempo de viagem quanto a distância percorrida.

As Tabelas 5.2 e 5.3 mostram um comparativo entre os dados de qualidade de vida das regionais de Belo Horizonte, especificamente, os valores do IDHM e suas três dimensões (Educação, Longevidade e Renda), e o valor da mediana do *Indicador de Ineficiência do Transporte Público* de cada regional. A Tabela 5.4 mostra os dados quando o indicador é calculado utilizando as rotas obtidas tendo os serviços de saúde de cada regional como destino, e a Tabela 5.5 apresenta os resultados utilizando as rotas criadas tendo os serviços de educação como destino.

Tabela 5.3: Tabela comparativa entre os dados de qualidade de vida (IDH) e o valor da mediana do *Indicador de Ineficiência do Transporte Público* por regional para os serviços de educação de Belo Horizonte.

Regional	IDH	IDH - Educação	IDH - Longevidade	IDH - Renda	Mediana - <i>Indicador de Ineficiência do Transporte Público</i>
Venda Nova	0.755	0.683	0.860	0.732	32.72
Leste	0.827	0.748	0.910	0.830	24.06
Norte	0.754	0.679	0.862	0.733	23.47
Barreiro	0.744	0.672	0.852	0.720	14.76
Centro-sul	0.914	0.823	0.933	0.995	11.43
Pampulha	0.853	0.793	0.911	0.859	10.80
Nordeste	0.801	0.719	0.895	0.800	10.66
Oeste	0.839	0.757	0.910	0.857	10.47
Noroeste	0.818	0.744	0.907	0.811	8.72

Tabela 5.4: Tabela comparativa entre o número de serviços na regional, as médias do tempo de viagem e da distância percorrida pelas rotas, e o valor da mediana do *Indicador de Ineficiência do Transporte Público* para os serviços de saúde de Belo Horizonte.

Regional	N° de serviços	$\bar{T}$ - Privado	$\bar{T}$ - Público	$\bar{D}$ - Privado	$\bar{D}$ - Público	Mediana Indicator
Venda Nova	36	6 min e 26s	49 min e 7s	2.67 km	9.65 km	27.39
Norte	39	8 min e 44s	50 min e 40s	3.58 km	13.08 km	19.5
Leste	39	6 min e 29s	32 min e 55s	2.67 km	8.99 km	16.52
Barreiro	46	7 min e 55s	32 min e 57s	3.28 km	8.66 km	14.32
Nordeste	36	9 min e 29s	40 min e 41s	3.97 km	10.66 km	13.00
Noroeste	36	8 min e 17s	32 min e 8s	3.44 km	8.81 km	10.59
Pampulha	35	10 min e 40s	49 min e 17s	4.49 km	11.41 km	10.42
Oeste	35	7 min e 14s	31 min e 39s	2.93 km	8.02 km	9.29
Centro-sul	46	7 min e 5s	32 min e 15s	2.84 km	6.84 km	8.15

Tanto a Tabela 5.4 quanto a Tabela 5.5 visam complementar os dados apresentados pelas Tabelas 5.2 e 5.3, permitindo comparar o valor da mediana do *Indicador de Ineficiência do Transporte Público* das regionais com o número de serviços da categoria escolhida, e as médias do tempo de viagem ( $\bar{T}$ ) e da distância percorrida ( $\bar{D}$ ) pelas rotas. A Tabela 5.4 apresenta esses dados para os serviços de saúde, e a Tabela 5.5 para os serviços de educação.

Como apresentado na Seção 4.2.3, o *Indicador de Ineficiência do Transporte Pú-*



Tabela 5.5: Tabela comparativa entre o número de serviços na regional, as médias do tempo de viagem ( $\bar{T}$ ) e da distância percorrida ( $\bar{D}$ ) pelas rotas, e o valor da mediana do *Indicador de Ineficiência do Transporte Público* para os serviços de educação de Belo Horizonte.

Regional	Nº de serviços	$\bar{T}$ - Privado	$\bar{T}$ - Público	$\bar{D}$ - Privado	$\bar{D}$ - Público	Mediana Indicator
Venda Nova	124	6 min e 13s	50 min e 15s	2.57 km	9.85 km	32.72
Leste	140	6 min e 32s	37 min e 38s	2.68 km	10.09 km	24.06
Norte	104	8 min e 5s	49 min e 49s	3.30 km	12.51 km	23.47
Barreiro	157	8 min e 11s	32 min e 46s	3.40 km	8.92 km	14.76
Centro-sul	202	6 min e 51s	35 min e 11s	2.73 km	7.41 km	11.43
Pampulha	148	10 min e 31s	46 min e 02s	4.44 km	10.54 km	10.80
Nordeste	150	9 min e 50s	43 min e 49s	4.09 km	11.33 km	10.66
Oeste	177	7 min e 55s	33 min e 17s	3.24 km	8.53 km	10.47
Noroeste	152	8 min e 11s	32 min e 53s	3.37 km	8.83 km	8.72

*blico* é calculado como o produto entre o quociente da distância percorrida pela rota gerada pelo transporte público pela distância percorrida pela rota criada para o transporte privado, e o quociente entre o tempo de viagem da rota gerada pelo transporte público e o tempo de viagem da rota criada pelo transporte privado. Assim, se cada um desses quocientes tem valor maior que 1, temos um indicativo que a rota gerada pelo sistema de transporte público é pior que a gerada pelo transporte privado. Como o indicador é um produto desses quocientes, um valor muito alto indica uma ineficiência do transporte público em relação ao transporte privado. Os resultados apresentados nas Tabelas 5.2 e 5.3 mostram que, para todas as regionais, e considerando as duas categorias de serviços, as rotas geradas pelo transporte público são muito piores que as geradas pelo transporte privado. Isso é corroborado pelos valores das médias de tempo de viagem e de distância percorrida apresentadas nas Tabelas 5.4 e 5.5, e vem complementar as análises feitas na Seção 5.1, que indicam que as rotas criadas para o transporte público são mais distantes da ótima que as criadas pelo transporte privado. Outra conclusão importante é a de que, de um modo geral, os valores dos indicadores de qualidade de vida e da mediana do *Indicador de Ineficiência do Transporte Público* de uma regional estão relacionados, isso avaliando tanto os dados de acesso aos serviços de saúde, quanto aos de educação. No entanto, tal relação nem sempre é direta, uma vez que, o IDH e suas componentes são calculados levando em consideração diversos fatores, sobre os quais o acesso a serviços básicos não está necessariamente ligado. De maneira alguma o *Indicador de Ineficiência do Transporte Público* foi criado com o

intuito de explicar as diferenças nos índices de qualidade de vida das regionais de uma cidade, mas sim para permitir realizar a correlação entre tais dados. Dito isso, algumas análises interessantes podem ser feitas a partir dos dados apresentados:

- Em alguns casos, o valor da mediana do *Indicador de Ineficiência do Transporte Público* de uma regional é alto, mas o valor absoluto das médias do tempo de viagem pelo sistema de transporte público e da distância percorrida por tal sistema, não são tão maiores se comparados com o de outras regionais. Esse é o caso, por exemplo, da regional Venda Nova, que apresenta o maior valor de mediana do indicador para ambas as categorias de serviços, sem ter necessariamente o maior valor absoluto para a distância percorrida ou tempo de viagem das rotas criadas pelo sistema de transporte público. Isso é razoável, uma vez que o indicador é calculado como o quociente entre as variáveis tempo e distância, e o que ele indica é o quanto o sistema de transporte público é ineficiente ao ser comparado com o sistema de transporte privado, considerando tempo de viagem e distância percorrida. Assim, ao verificarmos os valores das médias para a regional Venda Nova, percebemos que comparativamente, o sistema de transporte público tem um tempo de viagem mais longo e percorre uma distância maior que o privado, e essa diferença é de fato maior que a observada em outras regionais, como a Norte ou a Leste, apesar de tais regionais terem valores absolutos maiores para as médias.
- A regional Leste, para ambas as categorias de serviço de destino tem um valor alto da mediana do *Indicador de Ineficiência do Transporte Público*, apesar ter um IDH relativamente alto. Para o caso dessa regional, não apenas comparativamente as rotas produzidas pelo sistemas de transporte público tem maior tempo de viagem e percorrem uma maior distância, como os valores absolutos das médias para tais fatores são altos, sobretudo para a distância percorrida. Apesar disso, o número de serviços tanto de saúde quanto de educação nessa regional é relativamente alto se comparado com o de outras regionais com valores de mediana menores, ou seja, apesar da regional ter muitos serviços disponíveis, o acesso a tais serviços pelo transporte público é ruim, se comparado ao acesso usando o sistema privado.
- Outro fato interessante é o de que a regional Centro-Sul, é a que possui o maior número de serviços tanto de saúde, quanto de educação (o que é razoável já que é a regional mais central da cidade). No entanto, ela nem sempre apresenta o melhor acesso a tais serviços, como mostram os dados da Tabela 5.5. Os dados

dessa tabela indicam que, de fato, comparativamente, o acesso aos serviços de educação da regional Noroeste pelo transporte público é melhor que o da regional centro-sul.

- A regional Barreiro é a que possui os piores valores para os indicadores de qualidade de vida. No entanto, para ambas as categorias de serviços ela possui um número grande de unidades oferecendo o serviço para a população, médias de tempo de viagem e de distância percorrida por ambos os sistemas de transporte relativamente baixos e próximos dos demais regionais, e conseqüentemente um valor de mediana do *Indicador de Ineficiência do Transporte Público* relativamente baixo. Isso indica que a regional é bem assistida com relação aos serviços básicos de saúde e educação, e que o acesso a tais serviços pelo transporte público não é muito ruim se comparado ao acesso pelo sistema privado.

A análise dos dados apresentados permite aos planejadores urbanos a identificação de regionais da cidade que são mal atendidas tanto no número de serviços de saúde e de educação, quanto no acesso utilizando o sistema de transporte público (caso das regionais Venda Nova e Norte), regionais que são bem atendidas no número de serviços, mas tem o acesso pelo transporte público ruim (como a regional Leste), e regionais que mesmo sendo bem atendidas no número e no acesso aos serviços, tem índices de qualidade de vida ruins (caso da regional Barreiro). A identificação desses problemas permite a criação de políticas públicas que visem não só a melhoria da acessibilidade e da mobilidade urbana, mas também da melhoria da qualidade de vida do cidadão.



# Capítulo 6

## Conclusão

Neste trabalho foram estudadas técnicas e métodos que permitem a análise, integração e visualização de dados urbanos de múltiplas fontes heterogêneas visando a criação de um modelo genérico para análise de dados urbanos, com foco em dados de transporte e trânsito. O Capítulo 3 apresenta esse modelo, criado com o intuito de ser utilizado como um guia na construção de aplicações e sistemas em Computação Urbana. O modelo possui três camadas: Integração, Armazenamento de Dados e Prestação de serviços, cada uma responsável por um aspecto específico da aplicação ou sistema. A camada de Integração é responsável pela ingestão dos dados brutos de suas fontes, e pelo pré-processamento, integração e padronização desses dados. Na camada de armazenamento de dados, os dados já padronizados são então persistidos em alguma estrutura específica, em geral, em um SGBD. A camada de prestação de serviços implementa os algoritmos de visualização de dados, mineração de dados e aprendizado de máquina que são utilizados na aplicação, além de uma interface que permita o acesso do usuário ao serviço criado.

No Capítulo 4 apresentamos dois estudos de caso que utilizam o modelo criado como guia, o que comprova a viabilidade, aplicabilidade e validade de tal modelo na construção de aplicações em Computação Urbana. O uso do modelo facilitou a construção dos estudos de caso, permitindo criar um fluxo de construção e organizar os algoritmos e programas criados em camadas, com propósitos específicos, e análogas às definidas no modelo. Sendo assim, o modelo auxiliou na definição da arquitetura das soluções criadas.

O primeiro estudo de caso é a *Impressão Digital de Trânsito Urbano*, que foi inspirada no *Urban Mobility Fingerprint* criado por Reimann et al. [34]. A ideia principal da *Impressão Digital de Trânsito Urbano* é permitir a comparação do acesso e da mobilidade urbana em uma cidade utilizando as modalidades de transporte público (mais

especificamente os ônibus) e privado. Assim, precisamos ter acesso aos dados da rede de circulação ou rede viária da cidade, e aos dados da rede de itinerários de ônibus. No trabalho, nos concentramos na obtenção desses dados para as cidades de São Paulo e Belo Horizonte. A *Impressão Digital de Trânsito Urbano* permite tanto a comparação visual e qualitativa, através das visualizações criadas, quanto quantitativa, através dos valores do *Indicador de Ineficiência da Rota*, da otimalidade das rotas produzidas. O segundo estudo de caso é o *Mapa Urbano de Acesso a Serviços Básicos*, criado com o intuito de permitir a comparação da eficiência das rotas produzidas pelos sistema de transporte público e privado, além de contrastar esses dados com dados de qualidade de vida de uma cidade. A ideia principal do *Mapa Urbano de Acesso a Serviços Básicos* é permitir comparar o acesso a serviços básicos para a população como saúde (hospitais, postos de saúde, clínicas, farmácias, etc) e educação (escolas, faculdades, creches, etc), utilizando os diferentes sistemas de transporte, e comparar tais dados com dados de qualidade de vida das regionais. Logo, para a construção do *Mapa Urbano de Acesso a Serviços Básicos* precisamos: dos dados das redes viária e de itinerários de ônibus de uma cidade, das regionais da cidade, dos dados de qualidade de vida de cada regional, e dos serviços de saúde e de educação localizados em cada regional. Neste trabalho, nos concentramos na obtenção de tais dados para a cidade de Belo Horizonte.

No Capítulo 5 são apresentadas análises construídas utilizando as duas ferramentas criadas. De um modo geral, os resultados obtidos podem de fato auxiliar na tomada de decisões sobre planejamento urbano, mostrando áreas em que o acesso utilizando sobretudo o sistema de transporte público é ruim, tanto do ponto de vista da otimalidade das rotas criadas, quanto da eficiência dessas rotas ao serem comparadas com aquelas criadas para o sistema de transporte privado. É possível também identificar áreas em que o número de serviços de educação e de saúde é deficitário, e relacionar tais informações com dados de qualidade de vida. A identificação e análise desses problemas permite ao planejador urbano promover políticas públicas para a melhoria tanto da acessibilidade e da mobilidade urbana, quanto da qualidade de vida das pessoas. Fica claro a partir dos dados apresentados que o acesso nas cidades estudadas usando o sistema de transporte público é em geral muito pior se comparado ao transporte privado, o que mostra a situação de declínio no uso do transporte de massa, devido à combinação de baixa qualidade e alto custo. O custo do transporte público não pode compensar as desvantagens percebidas em termos de distância e tempo quando comparado ao transporte individual, e os resultados apresentados corroboram esse fato. Assim, a redução do uso do transporte público pode ter graves consequências para o sistema geral de trânsito e transporte, reduzindo a mobilidade e aumentando tanto os custos quanto a poluição. O uso do transporte de massa deve ser incentivado e

sua qualidade deve melhorar, mas considerando as análises aqui apresentadas, não há muito incentivo para os cidadãos persistirem em usar o transporte público, a menos que sejam forçados a fazê-lo, devido a limitações econômicas pessoais.

A partir da pesquisa proposta nessa dissertação, diversos trabalhos podem ser criados. Como o modelo proposto foi criado de forma genérica, ele pode ser o guia na criação de diversas aplicações e sistemas. Possíveis trabalhos futuros envolvem a criação de aplicativos que usam fontes de dados dinâmicas, como *streams* de GNSS e mídias sociais. Temos a possibilidade de criar variações interativas na *Impressão Digital de Trânsito Urbano*, com a escolha de outros esquemas cromáticos. Outra direção para trabalhos futuros é a integração de outros meios de transporte, como o sistema de metrô, bicicletas e o caminhar a pé nos dados já coletados, permitindo o desenvolvimento de análises de transporte multimodal. Considerações sobre custos também devem ser levadas em conta, para que um modelo de simulação mais completo, no qual políticas possam ser formuladas e testadas, possa ser projetado e construído. A avaliação da veracidade dos tempos calculados tanto para percorrer as arestas da rede viária, quanto as arestas da rede de itinerários de ônibus é outra fonte de trabalhos futuros, bem como a simulação do atendimento das linhas de ônibus, verificando os problemas de cobertura evidenciados no trabalho. Finalmente, a possibilidade de reavaliar os itinerários de ônibus para reduzir a discrepância de tempo nas rotas pode tornar os resultados produzidos pelas visualizações e indicadores mais próximos da situação real de trânsito de uma cidade.





# Referências Bibliográficas

- [1] Accorsi, P.; Lalande, N.; Fabrègue, M.; Braud, A.; Poncelet, P.; Sallaberry, A.; Bringay, S.; Teisseire, M.; Cernesson, F. & Ber, F. L. (2014). Hydroqual: Visual analysis of river water quality. Em *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 123–132. ISSN .
- [2] Andrienko, N. & Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3--24. ISSN 1473-8716.
- [3] Bach, B.; Shi, C.; Heulot, N.; Madhyastha, T.; Grabowski, T. & Dragicevic, P. (2016). Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568. ISSN 1077-2626.
- [4] Borkin, M.; Gajos, K.; Peters, A.; Mitsouras, D.; Melchionna, S.; Rybicki, F.; Feldman, C. & Pfister, H. (2011). Evaluation of artery visualizations for heart disease diagnosis. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2479–2488. ISSN 1077-2626.
- [5] Cao, N.; Shi, C.; Lin, W. S.; Lu, J.; Lin, Y.-R. & Lin, C.-Y. (2016). Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Trans. Vis. Comput. Graph.*, 22(1):280–289.
- [6] Carvalho, C. H. R. & Pereira, R. H. M. (2011). Efeitos da variação da tarifa e da renda da população sobre a demanda de transporte público coletivo urbano no brasil. Text for Discussion, Applied Economic Research Institute (IPEA) 1595.
- [7] Carvalho, C. H. R. & Pereira, R. H. M. (2012). Gastos das famílias brasileiras com transporte urbano público e privado no brasil: Uma análise da pof 2003 e 2009. Texto para Discussão, Instituto de Pesquisa Econômica Aplicada (IPEA) 1803, Brasília.

- [8] Castro, P. S.; Zhang, D.; Chen, C.; Li, S. & Pan, G. (2013). From taxi gps traces to social and community dynamics: A survey. volume 46, pp. 17:1--17:34, New York, NY, USA. ACM. ISSN 0360-0300.
- [9] Chung, K.-L.; Qu, H.; Chan, W.-Y.; Guo, P.; Xu, A. & Lau, K.-H. (2007). Visual analysis of the air pollution problem in hong kong. *IEEE Transactions on Visualization and Computer Graphics*, 13:1408–1415. ISSN 1077-2626.
- [10] Dietrich, C.; Koop, D.; Vo, H. T. & Silva, C. T. (2014). Baseball4d: A tool for baseball game reconstruction amp; visualization. Em *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 23–32. ISSN .
- [11] Duarte, F. S. L. G.; Sikansi, F.; Fatore, F. M.; Fadel, S. G. & Paulovich, F. V. (2014). Nmap: A novel neighborhood preservation space-filling algorithm. *IEEE Trans. Vis. Comput. Graph.*, 20(12):2063–2071.
- [12] Ferreira, N.; Poco, J.; Vo, H. T.; Freire, J. & Silva, C. T. (2013). Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149--2158. ISSN 1077-2626.
- [13] Gomide, A.; Pereira, R.; Henrique Ribeiro de Carvalho, C.; Ferreira Mation, L.; Balbim, R.; Lima Neto, V.; Galindo, E.; Krause, C. & Guedes, E. (2013). Tarifação e financiamento do transporte público urbano. Text for discussion, applied economic research institute (ipea), Brasília.
- [14] Goodwin, S.; Dykes, J.; Jones, S.; Dillingham, I.; Dove, G.; Duffy, A.; Kachkaev, A.; Slingsby, A. & Wood, J. (2013). Creative user-centered visualization design for energy analysts and modelers. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2516–2525. ISSN 1077-2626.
- [15] Guo, D. & Zhu, X. (2014). Origin-destination flow data smoothing and mapping. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2043–2052. ISSN 1077-2626.
- [16] Hurter, C.; Ersoy, O. & Telea, A. (2012). Graph bundling by kernel density estimation. *Comput. Graph. Forum*, 31(3pt1):865--874. ISSN 0167-7055.
- [17] Janetzko, H.; Sacha, D.; Stein, M.; Schreck, T.; Keim, D. A. & Deussen, O. (2014). Feature-driven visual analytics of soccer data. Em *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 13–22. ISSN .

- [18] Klemm, P.; Lawonn, K.; Glaßer, S.; Niemann, U.; Hegenscheid, K.; Völzke, H. & Preim, B. (2016). 3d regression heat map analysis of population study data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):81–90. ISSN 1077-2626.
- [19] Kloeckl, K.; Senn, O. & Ratti, C. (2012). Enabling the real-time city: Live singapore! *J. Urban Technol*, 2(19):89–112.
- [20] Kostakos, V. & O’Neill, E. (2008). Cityware: Urban computing to bridge online and real-world social networks. Em Foth, M., editor, *Handbook of research on urban informatics: The practice and promise of the real-time city*, pp. 195–204. IGI Global.
- [21] Krueger, R.; Thom, D. & Ertl, T. (2014). Visual analysis of movement behavior using web data for context enrichment. Em *2014 IEEE Pacific Visualization Symposium*, pp. 193–200. ISSN 2165-8765.
- [22] Lathia, N.; Froehlich, J. & Capra, L. (2010). Mining public transport usage for personalised intelligent transport systems. *10th IEEE International Conference on Data Mining*, p. 887–892.
- [23] Li, J.; Zhang, K. & Meng, Z.-P. (2014). Vismate: Interactive visual analysis of station-based observation data on climate changes. *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 00:133–142. ISSN .
- [24] Liu, H.; Gao, Y.; Lu, L.; Liu, S.; Qu, H. & Ni, L. M. (2011a). Visual analysis of route diversity. Em *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 171–180. ISSN .
- [25] Liu, S.; Cui, W.; Wu, Y. & Liu, M. (2014). A survey on information visualization: Recent advances and challenges. *Vis. Comput.*, 30(12):1373–1393. ISSN 0178-2789.
- [26] Liu, S.; Maljovec, D.; Wang, B.; Bremer, P.-T. & Pascucci, V. (2015). Visualizing High-Dimensional Data: Advances in the Past Decade. Em Borgo, R.; Ganovelli, F. & Viola, I., editores, *Eurographics Conference on Visualization (EuroVis) - STARs*. The Eurographics Association.
- [27] Liu, Y.; Zheng, Y.; Yuan, J. & Xie, X. (2011b). Urban computing with taxicabs. *ACM - 13th International Conference on Ubiquitous Computing*, pp. 89–98.
- [28] Munzner, T. (2015). *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press. ISBN 9781498759717.

- [29] Mühlbacher, T. & Piringer, H. (2013). A partition-based framework for building and validating regression models. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):1962–1971. ISSN 1077-2626.
- [30] Pan, B.; Zheng, Y.; Wilkie, D. & Shahabi, C. (2013). Crowd sensing of traffic anomalies based on human mobility and social media. Em *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'13, pp. 344–353, New York, NY, USA. ACM.
- [31] Paulos, E. & Goodman, E. (2004). The familiar stranger: anxiety, comfort, and play in public places. Em Dykstra-Erickson, E. & Tscheligi, M., editores, *CHI*, pp. 223–230. ACM.
- [32] Psyllidis, A.; Bozzon, A.; Bocconi, S.; & Bolivar, C. T. (2015). A platform for urban analytics and semantic data integration in city planning. *CAAD Futures 2015*, (527):21–36.
- [33] Ratti, C.; Sobolevsky, S.; Calabrese, F.; Andris, C.; Reades, J.; Martino, M.; Claxton, R. & Strogatz, S. H. (2010). Redrawing the map of great britain from a network of humani. *PLoS ONE*, 12(5).
- [34] Reimann, R.; Groß, B. & Schmitt, P. (2017). All roads to rome: Visualizing mobility at scale. Em *2017 IEEE VIS Arts Program (VISAP)*, pp. 1–4. ISSN .
- [35] Robertson, G.; Fernandez, R.; Fisher, D.; Lee, B. & Stasko, J. (2008). Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1325--1332. ISSN 1077-2626.
- [36] Shi, C.; Fu, S.; Chen, Q. & Qu, H. (2015). Vismoc: Visualizing video clicks-stream data from massive open online courses. Em *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 159–166. ISSN 2165-8765.
- [37] Shklovski, I. & Chang, M. F. (2006). Guest editors' introduction: Urban computing—navigating space and context. volume 39, pp. 36–37, Los Alamitos, CA, USA. IEEE Computer Society Press. ISSN 0018-9162.
- [38] Speckmann, B.; Verbeek, K. & Buchin, K. (2011). Flow map layout via spiral trees. *IEEE Transactions on Visualization & Computer Graphics*, 17:2536–2544. ISSN 1077-2626.

- [39] Thiagarajan, A.; Ravindranath, L.; LaCurts, K.; Madden, S.; Balakrishnan, H.; Toledo, S. & Eriksson, J. (2009). Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones. *7th ACM Conference on Embedded Networked Sensor Systems*.
- [40] Tominski, C.; Schumann, H.; Andrienko, G. & Andrienko, N. (2012). Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2565–2574. ISSN 1077-2626.
- [41] Trimm, D.; Rheingans, P. & desJardins, M. (2012). Visualizing student histories using clustering and composition. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2809–2818. ISSN 1077-2626.
- [42] Wang, Z.; Lu, M.; Yuan, X.; Zhang, J. & v. d. Wetering, H. (2013). Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2159–2168. ISSN 1077-2626.
- [43] Wu, W.; Xu, J.; Zeng, H.; Zheng, Y.; Qu, H.; Ni, B.; Yuan, M. & Ni, L. M. (2016). Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data. *IEEE Trans. Vis. Comput. Graph.*, 22(1):935–944.
- [44] Wu, W.; Zheng, Y.; Qu, H.; Chen, W.; Gröller, E. & Ni, L. M. (2014). Boundary-seer: Visual analysis of 2d boundary changes. Em *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 143–152. ISSN .
- [45] Xie, C.; Chen, W.; Huang, X.; Hu, Y.; Barlowe, S. & Yang, J. (2014). Vaet: A visual analytics approach for e-transactions time-series. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1743–1752. ISSN 1077-2626.
- [46] Yuan, J.; Zheng, Y. & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and pois. *ACM - 18th SIGKDD Conference on Knowledge Discovery and Data Mining*, p. 186–194.
- [47] Zhang, J.; Yanli, E.; Ma, J.; Zhao, Y.; Xu, B.; Sun, L.; Chen, J. & Yuan, X. (2014). Visual analysis of public utility service problems in a metropolis. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1843–1852. ISSN 1077-2626.
- [48] Zhao, J.; Cao, N.; Wen, Z.; Song, Y.; Lin, Y. & Collins, C. (2014). #fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE Trans. Vis. Comput. Graph.*, 20(12):1773–1782.

- [49] Zheng, Y.; Capra, L.; Wolfson, O. & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. ACM TIST.
- [50] Zheng, Y.; Liu, Y.; Yuan, J. & Xie, X. (2011). Urban computing with taxicabs. Em *Proceedings of the 13th International Conference on Ubiquitous Computing, UbiComp '11*, pp. 89--98, New York, NY, USA. ACM.
- [51] Zheng, Y.; Wu, W.; Chen, Y.; Qu, H. & Ni, L. M. (2016). Visual analytics in urban computing: An overview. *IEEE Transactions on Big Data*, 2(3):276–296.
- [52] Zheng, Y.; Zhang, L.; Xie, X. & Ma, W. Y. (2009). Mining interesting locations and travel sequences from gps trajectories. *18th International Conference on World Wide Web*, p. 791–800.
- [53] Zhong, C.; Arisona, S. M. & Schmitt, G. (2014). A visual analytics framework for large transportation dataset. *19th International Conference on Computer-Aided Architectural Design Research in Asia CAADRIA*, 5(19).
- [54] Ziegler, H.; Jenny, M.; Gruse, T. & Keim, D. A. (2010). Visual market sector analysis for financial time series data. Em *2010 IEEE Symposium on Visual Analytics Science and Technology*, pp. 83–90. ISSN .
- [55] Zimmerman, J.; Tomasic, A.; Garrod, C.; Yoo, D.; Hiruncharoenvate, C.; Aziz, R.; Thiruvengadam, N. R.; Huang, Y. & Steinfeld, A. (2011). Field trial of tiramisu: Crowd-sourcing bus arrival times to spur co-design. *ACM - Annual Conference on Human Factors in Computing Systems*, p. 1677–1686.

# Anexo A

## Dados Utilizados

Os dados preparados e utilizados nos estudos de caso apresentados nesta dissertação estão disponíveis para acesso interativo e download na Infraestrutura de Dados Espaciais (IDE) do Laboratório CS+x do Departamento de Ciência da Computação da UFMG. Os links para acesso são os seguintes:

- Rede de Itinerários de Ônibus de São Paulo:
  - Nós:  
[http://aqui.io/geodata/layers/opengeo%3Anos\\_rede\\_onibus\\_sp](http://aqui.io/geodata/layers/opengeo%3Anos_rede_onibus_sp)
  - Arestas:  
[http://aqui.io/geodata/layers/opengeo%3Aedges\\_rede\\_onibus\\_sp](http://aqui.io/geodata/layers/opengeo%3Aedges_rede_onibus_sp)
  
- Rede de Itinerários de Ônibus de Belo Horizonte:
  - Nós:  
[http://aqui.io/geodata/layers/opengeo%3Anos\\_rede\\_onibus\\_bh](http://aqui.io/geodata/layers/opengeo%3Anos_rede_onibus_bh)
  - Arestas:  
[http://aqui.io/geodata/layers/opengeo%3Arede\\_itinerarios\\_onibus\\_edges\\_1](http://aqui.io/geodata/layers/opengeo%3Arede_itinerarios_onibus_edges_1)
  
- Rede Viária de São Paulo:
  - Nós:  
[http://aqui.io/geodata/layers/opengeo%3Anos\\_rede\\_viaria\\_sp](http://aqui.io/geodata/layers/opengeo%3Anos_rede_viaria_sp)
  - Arestas:  
[http://aqui.io/geodata/layers/opengeo%3Aarestas\\_rede\\_viaria\\_sp](http://aqui.io/geodata/layers/opengeo%3Aarestas_rede_viaria_sp)

- Rede Viária de Belo Horizonte:
  - Nós:  
[http://aqui.io/geodata/layers/opengeo%3Abh\\_road\\_nodes\\_from\\_osm\\_2017\\_04](http://aqui.io/geodata/layers/opengeo%3Abh_road_nodes_from_osm_2017_04)
  - Arestas:  
[http://aqui.io/geodata/layers/opengeo%3Abh\\_roads\\_from\\_osm\\_2017\\_04](http://aqui.io/geodata/layers/opengeo%3Abh_roads_from_osm_2017_04)
- Serviços de Saúde de Belo Horizonte:  
[http://aqui.io/geodata/layers/opengeo%3Aservicos\\_saude\\_bh](http://aqui.io/geodata/layers/opengeo%3Aservicos_saude_bh)
- Serviços de Educação de Belo Horizonte:  
[http://aqui.io/geodata/layers/opengeo%3Aservicos\\_educacao\\_bh](http://aqui.io/geodata/layers/opengeo%3Aservicos_educacao_bh)
- Regionais de Belo Horizonte:  
<http://aqui.io/geodata/layers/belohorizonte%3Aregional>