



ANÁLISE DA CAPACIDADE DE DESCARREGAMENTO DE DADOS USANDO O SISTEMA DE TRANSPORTE PÚBLICO DE UMA METRÓPOLE

Daniel de Souza Dias

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Luís Henrique Maciel Kosmalski
Costa

Rio de Janeiro
Fevereiro de 2018

ANÁLISE DA CAPACIDADE DE DESCARREGAMENTO DE DADOS
USANDO O SISTEMA DE TRANSPORTE PÚBLICO DE UMA METRÓPOLE

Daniel de Souza Dias

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Luís Henrique Maciel Kosmalski Costa, Dr.

Prof. Célio Vinicius Neves de Albuquerque, Ph.D.

Prof. Marcelo Gonçalves Rubinstein, D.Sc.

Prof. Mauro Sergio Pereira Fonseca, Dr.

Prof. Miguel Elias Mitre Campista, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
FEVEREIRO DE 2018

Dias, Daniel de Souza

Análise da capacidade de descarregamento de dados usando o sistema de transporte público de uma metrópole/Daniel de Souza Dias. – Rio de Janeiro: UFRJ/COPPE, 2018.

XIII, 109 p.: il.; 29, 7cm.

Orientador: Luís Henrique Maciel Kosmalski Costa

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2018.

Referências Bibliográficas: p. 95 – 102.

1. Big-Data. 2. Redes veiculares. 3. Dados de GPS dos ônibus da cidade do Rio de Janeiro. 4. Transferência de dados. I. Costa, Luís Henrique Maciel Kosmalski. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*Love and work... Work and love,
that's all there is.
Sigmund Freud*

Agradecimentos

À minha família pelo exemplo, apoio, compreensão, paciência e fortaleza nos momentos mais difíceis da minha vida. Em especial, aos meus exemplos de vida, Sérgio e Luciane, que não mediram esforços para que nada nos faltasse e não mediram esforços para que eu chegasse até esta etapa da minha vida.

À minha querida esposa Kamilla, que sempre consegue arrancar-me um sorriso e me presenteou com inúmeros momentos felizes.

Ao meu orientador, Luís Henrique, pela paciência na orientação e pelo incentivo que tornou possível a conclusão deste trabalho.

Agradeço aos professores Célio Vinicius Neves de Albuquerque, Igor Moraes, Marcelo Gonçalves Rubinstein, Mauro Sergio Pereira Fonseca e Miguel Campista pelas sugestões na qualificação e pela participação na banca examinadora desta tese.

A todos os professores do GTA/UFRJ, Luís Henrique, Aloysio Pedroza, Otto Duarte, Miguel Campista e Pedro Velloso, que com sua dedicação e competência, fazem deste laboratório um centro de excelência em pesquisa e inovação.

Aos amigos e colegas que fiz no GTA, pela ajuda no meu crescimento pessoal e profissional, pelo incentivo, apoio e bom humor constantes.

Ao Guilherme Mori, da empresa 99, por ter se interessado pelo trabalho e gentilmente colaborado com a pesquisa.

À prefeitura da cidade do Rio de Janeiro, por disponibilizar os dados de localizações dos ônibus ao público.

À Universidade Federal do Rio de Janeiro (UFRJ) e a todos os professores da COPPE, na qual me orgulho de ter vivido grandiosos momentos desde a Graduação.

Agradeço aos funcionários do Programa de Engenharia Elétrica da COPPE/UFRJ, Maurício, Daniele, Roberto e Marcos pela presteza no atendimento na secretaria do Programa.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

ANÁLISE DA CAPACIDADE DE DESCARREGAMENTO DE DADOS USANDO O SISTEMA DE TRANSPORTE PÚBLICO DE UMA METRÓPOLE

Daniel de Souza Dias

Fevereiro/2018

Orientador: Luís Henrique Maciel Kosmowski Costa

Programa: Engenharia Elétrica

Cidades inteligentes melhorarão a qualidade de vida de seus cidadãos com a ajuda das tecnologias da informação e da comunicação. Entre outras tarefas, é necessário o monitoramento ambiental e a análise das informações para determinar ações apropriadas. A realização dessas tarefas depende de infraestrutura de comunicação confiável e ubíqua, como uma rede celular. No entanto, a rede celular pode não ser capaz de lidar com as necessidades de pico, deixando de prestar serviços quando a cidade mais precisa. Uma solução promissora, objeto deste estudo, é utilizar canais de comunicação alternativos para aliviar a infraestrutura da rede celular. Nesta tese investiga-se a transmissão de dados usando redes veiculares como um canal de descarregamento de dados. Mais especificamente, é analisada uma rede veicular formada por veículos de transporte público, ônibus e táxis, que circulam pela cidade ao longo de todo o dia. Utilizar a onipresença de ônibus e táxis para ajudar a reunir e mover dados, parece promissor; entretanto, deve-se garantir que eles possam fornecer coletivamente a dinâmica e a capacidade necessárias para satisfazer os requisitos de transmissão. Analisa-se também a dinâmica e a capacidade da rede veicular considerando diferentes tecnologias sem-fio atualmente disponíveis. Dada a escala da rede e o grande número de interações entre veículos, inicialmente foi necessário reduzir a complexidade do sistema adotando uma topologia de grade e agrupando em clusters veículos próximos usando o algoritmo STING. As análises revelam aspectos interessantes, como o fato dos veículos apresentarem diferentes padrões de movimentos. Além disso, ônibus possuem um tempo médio inferior ao dos táxis para alcançar o próximo ponto de transferência. Finalmente, os resultados mostram que quantidades significativas de dados podem ser transportadas na cidade do Rio de Janeiro, da ordem de 500 TB por dia.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

ANALYSIS OF DATA OFFLOADING CAPACITY USING THE PUBLIC TRANSPORT SYSTEM OF A METROPOLIS

Daniel de Souza Dias

February/2018

Advisor: Luís Henrique Maciel Kosmowski Costa

Department: Electrical Engineering

Smart cities will improve the quality of life of their citizens with the help of information and communication technologies. Among other tasks, it involves environmental monitoring and information analysis to determine appropriate actions. The accomplishment of these tasks depends on a reliable, ubiquitous communication infrastructure, such as a cellular network. Nevertheless, the cellular network may not be sufficient to handle peak loads and may fail to provide services when the city needs it most. A promising solution, object of this study, is to rely on alternative communication channels to alleviate the cellular infrastructure. This thesis investigates the transmission of data using vehicular networks as a data offloading channel. More specifically, it is analyzed a vehicular network of public transport vehicles, composed by buses and taxis, which circulate at the city throughout the day. Relying on the ubiquitous presence of buses and taxis to help gather and move data sounds promising, nevertheless we must make sure that they can collectively provide the necessary dynamics and capacity to satisfy the transmissions requirements. It is analyzed the dynamics and capacity of the vehicular network by considering different wireless technologies currently available. Given the scale of the network and the huge number of interactions between vehicles, we first reduce the complexity of the system by adopting a grid topology and grouping in clusters nearby nodes using the STING algorithm. Our analyses reveal aspects like vehicles having different patterns of movements. Also, buses have a lower mean time than taxis to reach the next transfer point. Finally, the results show that significant amounts of data can be offloaded onto the vehicular network in the city of Rio de Janeiro, reaching 500 TB daily.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
1.1 Objetivos	2
1.2 Organização do Trabalho	4
2 Estado da Arte	6
2.1 Transporte de dados por veículos	6
2.1.1 Transmissões de dados por quaisquer veículos	7
2.1.2 Veículos de transporte público urbanos	9
2.1.3 Tecnologias de transmissão sem-fio	12
2.1.4 Bancos de dados de deslocamento disponíveis	13
2.2 Aplicações	14
2.2.1 Cidades inteligentes	14
2.2.2 Distribuição de informações para veículos	15
2.2.3 Métodos de transporte de grandes dados	16
3 Coleta dos Dados de Veículos Públicos Coletivos da Cidade do Rio de Janeiro	18
3.1 Infraestrutura de transporte na cidade do Rio de Janeiro	18
3.2 Dados dos ônibus urbanos da cidade do Rio de Janeiro	19
3.2.1 Coleta de dados de localização dos ônibus	21
3.2.2 Dados de passageiros	22
3.2.3 Identificação dos erros de localização	22
3.2.4 Análise dos padrões de deslocamento dos ônibus	26
3.3 Dados de posicionamento dos táxis	33
3.3.1 Análise dos padrões de deslocamento dos táxis	34
3.3.2 Análise dos padrões de deslocamento dos veículos públicos da cidade do Rio de Janeiro	36
3.4 Conclusão	39

4	Metodologia	41
4.1	Análise quantitativa dos dados	42
4.1.1	Análise preliminar	43
4.1.1.1	Escolha de parâmetros para a análise	44
4.1.1.2	Método de análise	45
4.1.1.3	Dados obtidos	46
4.1.1.3.1	Quantidade de contatos	46
4.1.1.3.2	Média de contatos	47
4.1.1.3.3	Características dos contatos entre ônibus	48
4.1.2	Tempo de execução da análise preliminar	50
4.1.3	Resultados da análise preliminar	51
4.2	Ferramentas de análise	51
4.2.1	Algoritmo de clusterização STING	52
4.3	Métodos de transmissão de dados	54
4.3.1	Tecnologias de transmissão	54
4.3.2	Rede modelada	55
4.4	Cálculo de deslocamento dos veículos	57
4.4.1	Algoritmo de cálculo de viagens	58
4.4.2	Tempo de execução do algoritmo de cálculo de viagens	60
4.4.3	Quantidade de dados de localização utilizados	61
4.5	Conclusão	62
5	Análise Estrutural	64
5.1	Dinâmica da rede	64
5.1.1	Número de atualizações por cluster	65
5.1.2	Mapa dos clusters	66
5.1.3	Distância entre pares de clusters	67
5.1.4	Viagens entre pares de clusters	71
5.1.5	Tempo médio de deslocamento entre pares de clusters	72
5.1.6	Tempo médio de permanência nos clusters	76
5.2	Análise da rede complexa	77
5.2.1	Viagens entre pares de clusters	77
5.3	Conclusão	79
6	Análise de Capacidade	81
6.1	Capacidade de transferência de dados	81
6.2	Latência de envio	86
6.3	Conclusão	88
7	Conclusões e Trabalhos Futuros	90

Referências Bibliográficas	95
A Tabelas com os resultados da análise preliminar	103
A.1 Dados de Contatos de um Ônibus	103
A.1.1 Quantidade de Contatos	103
A.1.2 Quantidade de ônibus únicos contactados	103
A.1.3 Média de contatos	104
A.1.4 Número de veículos contatados no ponto de contatos máximos	106
A.1.5 Linhas distintas	106
A.1.6 Média de tempo dos contatos	107
A.1.7 Contatos múltiplos	108
A.1.8 Velocidade média de contato	108

Lista de Figuras

1.1	Mapa com ambas as localizações no dia 07/10/2014.	3
3.1	Mapa com a localização dos ônibus no dia 07/10/2014.	22
3.2	Atualizações dos ônibus durante 24 horas do dia 07/10/2014.	27
3.3	Atualizações dos ônibus durante a semana de 06 a 12/10/2014.	28
3.4	Frequência de atualizações diárias por ônibus.	28
3.5	Tempo médio de intervalo de atualização dos ônibus.	29
3.6	Velocidade reportada pelo GPS dos ônibus.	30
3.7	Detalhamento da velocidade reportada pelo GPS dos ônibus.	30
3.8	Mapa com a velocidade dos ônibus entre 07:00h e 07:05h.	31
3.9	Número de linhas de um ônibus durante um dia.	32
3.10	Número de ônibus distintos que estão designados à linha.	32
3.11	Mapa com a localização dos táxis no dia 07/10/2014.	34
3.12	Atualizações dos táxis durante as 24 horas do dia 07/10/2014.	35
3.13	Frequência de atualizações diárias por táxis.	35
3.14	Tempo médio de intervalo de atualização dos táxis.	36
3.15	Mapa com localizações de ônibus e táxis as localizações no dia 07/10/2014.	37
3.16	Atualizações de ônibus e táxis durante as 24 horas do dia 07/10/2014.	38
3.17	Frequências de atualizações diárias.	38
3.18	Tempo médio de intervalo de atualização de ônibus e táxis simulta- neamente.	39
4.1	Número de atualizações do ônibus C302810 no banco de dados no dia 07/10/2014.	42
4.2	Mapa do Rio de Janeiro exibindo os 100 maiores clusters de 300 m. . .	54
4.3	Representação de viagens entre dois clusters.	60
4.4	Quantidade de dados de localização analisados pelo algoritmo.	62
4.5	Quantidade de veículos analisados pelo algoritmo.	62
5.1	Quantidade de atualizações por cluster.	65
5.2	Mapa com células de 300 m e 100 maiores clusters formados por ônibus.	66

5.3	Mapa com células de 300 m e 100 maiores clusters formados por táxis.	66
5.4	Mapa com células de 300 m e 100 maiores clusters formados por ônibus e táxis.	67
5.5	Mapa com a densidade demográfica da Cidade do Rio de Janeiro. . .	68
5.6	Distância (km) entre clusters de todos os veículos.	69
5.7	Distância (km) entre os 20 maiores clusters de todos os veículos. . . .	70
5.8	Histograma da distância (km) entre os clusters.	70
5.9	PDF do número de viagens entre pares de clusters.	71
5.10	Número de viagens por ID do cluster.	72
5.11	Número de viagens em relação à distância (km).	73
5.12	PDF do tempo médio de deslocamento entre pares de clusters.	73
5.13	Matriz do tempo médio de deslocamento entre os 20 primeiros clusters de ambos os veículos.	74
5.14	Tempo médio das viagens entre clusters em relação à distância (km).	75
5.15	Tempo médio das viagens de ambos os veículos entre clusters em relação à distância (km).	75
5.16	PDF do tempo de permanência em cada clusters.	77
5.17	Grafo representativo do número de viagens entre clusters de 300 metros.	78
6.1	Quantidade de veículos que reportam a sua localização em cada cluster.	82
6.2	Tempo total de permanência dos veículos nos clusters, por tipo de veículo.	83
6.3	Tempo total de permanência amostrado pelo período do dia.	84
6.4	Tempo total de permanência por tipo de veículo com clusters de 300 m.	85
6.5	CDF do tempo que os veículos levam para alcançar o próximo cluster.	86
6.6	Tempo médio gasto para encontrar o próximo cluster (tempo off-line).	87
6.7	Previsibilidade de encontrar um cluster.	88

Lista de Tabelas

3.1	Dados operacionais mensais da cidade do Rio de Janeiro (2014).	19
3.2	Metadados associados ao dataset – GPS ônibus.	20
3.3	Quantidade de viagens mensal - Outubro de 2014.	23
3.4	Erro de GPS do ônibus D867990.	23
3.5	Erro de GPS do ônibus A413540.	23
3.6	Erro de GPS do ônibus A410110.	24
3.7	Erro de GPS do ônibus B710430.	24
3.8	Erro de GPS do ônibus C414150.	25
3.9	Erro de GPS do ônibus A411830.	25
3.10	Erro de GPS do ônibus A480840.	26
4.1	Escolha dos ônibus analisados.	44
4.2	Quantidade de contatos.	47
4.3	Média de contatos por ponto.	48
4.4	Tempo médio dos contatos (segundos).	49
4.5	Tempo de execução da análise preliminar.	50
4.6	Visão geral das aplicações M2M.	55
4.7	Tecnologias de rede sem-fio consideradas.	56
4.8	Tempo de execução do Algoritmo 1.	61
6.1	Potencial de transferência de dados por dia (Terabytes).	84
A.1	Quantidade de contatos.	104
A.2	Quantidade de ônibus únicos contactados.	104
A.3	Porcentagem de ônibus únicos contactados.	105
A.4	Média de contatos.	105
A.5	Número de veículos contactados no ponto de contatos máximos.	106
A.6	Número de linhas únicas encontradas.	107
A.7	Média de tempo dos contatos (segundos).	107
A.8	Quantidade de contatos múltiplos.	108
A.9	Proporção de contatos múltiplos.	109
A.10	Velocidade média de contato (km/h).	109

Capítulo 1

Introdução

Cidades inteligentes prometem melhorar a qualidade de vida dos seus cidadãos através dos avanços das tecnologias de informação e de comunicação, permitindo ter conhecimento sobre as diversas condições na cidade de forma rápida. O conceito de cidade inteligente envolve monitorar variáveis do ambiente de uso comum e, por meio de algoritmos de análise especializados [1, 2] (*big data*), utilizar uma grande quantidade de informações para identificar e acionar corretamente os mecanismos apropriados para as ações necessárias ou simplesmente oferecer informações aos cidadãos da melhor maneira possível [3].

Ao aliar o transporte público, que inclui ônibus e táxis, com tecnologias de transmissões de dados, é possível utiliza-los em uma cidade inteligente em diversas funções [3], entre elas: métricas automatizadas de infraestrutura, tarifação diferenciada para serviços prestados pagos, como medição remota do consumo de gás ou eletricidade, sistema inteligente de drenagem e captação de dejetos, programas de monitoramento pluviométrico, da qualidade do ar e condições climáticas, sensoria-mento de neve com gerenciamento inteligente de limpeza e controle do sistema de transporte.

A atuação de forma correta na cidade depende fortemente da disponibilização de informações confiáveis, que devem ser obtidas por uma estrutura de comunicação capaz de abranger toda a área da cidade, como, por exemplo, a rede de telefonia celular. Contudo, a rede de telefonia celular, pode não ser suficiente para atender todas as necessidades de transmissão de informações, principalmente em situações de alta demanda e durante picos de uso. Estes momentos estão relacionados a situações críticas, como grandes eventos, grande fluxo de pessoas ou até mesmo situações de emergência que levem à indisponibilidade de recursos, como apagões e falta de telefonia fixa, ou seja, em momentos nos quais a cidade carece de informações para a tomada de decisões.

A utilização de um canal alternativo para a transmissão, que seja independente da rede de telefonia celular, é desejável. Entre as diversas formas disponíveis, uma

das possíveis soluções para minimizar o problema é a utilização de redes veiculares como meio de comunicação para diminuir a utilização da rede de telefonia celular – tópico abordado nesta tese.

As redes veiculares são redes móveis capazes de realizar comunicação de veículo para veículo (V2V) e de veículo para infraestrutura (V2I), onde essa infraestrutura normalmente é instalada no entorno das vias, e a comunicação ocorre através de múltiplas tecnologias de transmissão sem-fio. Esse tipo de rede é aplicado em 3 áreas principais, sendo elas: segurança no trânsito, entretenimento e assistência ao motorista [4]. Além disso, o papel das redes veiculares tende a crescer a medida que as tecnologias de transmissão evoluem e a sua penetração nos veículos é ampliada. A capacidade ociosa dessas redes pode ser aproveitada para outros fins, como, por exemplo, para a conexão dos veículos com dispositivos inteligentes diversos [5] ou como canal alternativo para a disseminação de mensagens.

1.1 Objetivos

Nesta tese é investigada a transmissão de dados usando redes veiculares como canal de descarregamento de dados. Através do contato entre dispositivos de rádio instalados em veículos e dispositivos de rádio instalados em pontos estratégicos distribuídos pela cidade, é possível utilizar essa infraestrutura para coletar dados que estão distribuídos pela cidade sem depender de outras redes, aliviando a quantidade de dados transmitidos na rede celular.

As informações a serem transmitidas podem ser obtidas por intermédio de contatos oportunistas entre veículos e dispositivos de monitoramento espalhados pela cidade, ou gerada por sensores instalados nos veículos. Os dados são armazenados nos veículos até que em um momento oportuno, quando estes estão localizados em um ponto estratégico contendo o dispositivo de rádio, são transmitidos para um local central. Dessa forma é possível a utilização de comunicação de curto alcance para coletar dados e para centralizá-los. Assim, a rede veicular é capaz de reduzir a quantidade de dados que são transmitidos através de redes que seriam especificamente criadas para a transmissão dessas informações ou de redes de telefonia celular, ambas com alto custo de implementação.

Utilizar os veículos de transporte público, contando com a ampla presença de ônibus e táxis para ajudar a disseminar os dados, oferece algumas vantagens. Esses veículos conseguem alcançar uma vasta área da cidade e, tipicamente, transitam durante o dia inteiro ao contrário de veículos particulares. Entretanto, algumas considerações devem ser analisadas. Somente na cidade do Rio de Janeiro, no ano de 2016, foram transportados aproximadamente de 1 bilhão e 300 milhões de passageiros, mais de 100 milhões mensalmente, em 1,6 milhões de viagens mensais em

mais de 8 mil ônibus [6]. Também circulam na cidade 32 mil táxis, de acordo com dados de abril de 2016.

Assim, é importante verificar se os veículos, de forma coletiva, estão aptos a fornecer o dinamismo e a capacidade necessários para as aplicações que se deseja implementar. A partir disso, levantam-se nesta tese duas questões:

- Como as regiões da cidade estão interligadas em relação ao deslocamento dos veículos?
- Quanta informação é possível transmitir entre essas regiões em função do tempo?

Diversas pesquisas estão relacionadas a essas questões, as quais são abordadas no Capítulo 2, que trata dos trabalhos relacionados. Esta tese contribui para o esforço da comunidade ao estudar a comunicação de dados utilizando redes veiculares a partir de uma grande quantidade de dados reais de localização de veículos de transporte públicos (ônibus e táxis) da cidade do Rio de Janeiro. A Figura 1.1 apresenta o mapa da cidade do Rio de Janeiro, na qual a localização de cada ônibus está representada como pontos pretos e a localização de cada táxi está representada como pontos vermelhos. Os dados de localização são do dia 7 de outubro de 2014 e correspondem ao período de 24h deste dia.

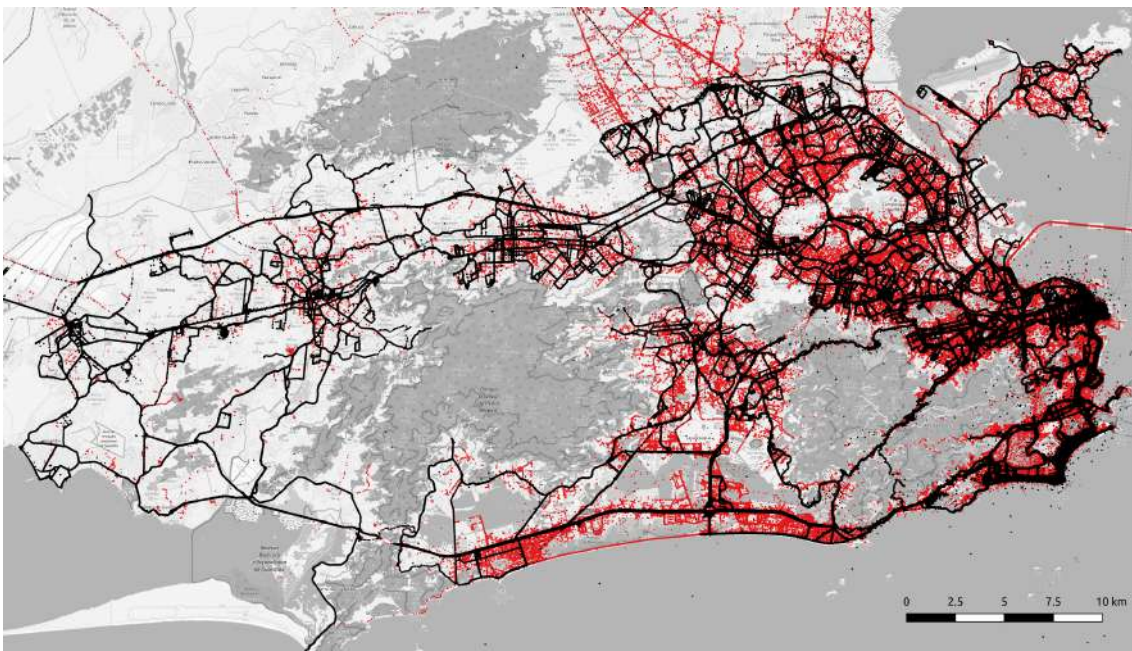


Figura 1.1: Mapa com ambas as localizações no dia 07/10/2014.

Nesta tese são examinados o dinamismo e a capacidade dessa rede veicular assumindo que cada veículo participante da rede possui um dispositivo com uma interface de rádio e que um segundo tipo de dispositivo, também com uma interface de

rádio, está localizado em diversos pontos estratégicos da cidade. Esses dispositivos são capazes de se comunicar com os veículos e centralizar os dados recebidos. Nesse cenário são analisadas distintas tecnologias de transmissão de dados sem-fio de curto alcance entre o primeiro e segundo dispositivos.

Não é trivial a determinação de todos os possíveis encontros entre todos os veículos dados a escala da rede devido a quantidade de veículos, o tamanho da região analisada e a imensa quantidade de interações entre veículos. O problema de descobrir todos os encontros dos veículos demanda ferramentas para trabalhar com grandes dados, como demonstrado no Capítulo 4.

Dessa forma, para reduzir a complexidade do problema e permitir executar a análise em tempo factível, foi utilizado um algoritmo de clusterização baseado em grade, ou *grid*, para agrupar os veículos relacionados. O algoritmo utilizado é o STING (*Statistical INformation Grid*), que consiste em dividir em grades, formadas por células geométricas idênticas, as coordenadas da cidade do Rio de Janeiro. Em cada célula são contabilizados o total de veículos e outras características intrínsecas, como quantidade de viagens de veículos entre clusters. O Capítulo 4 define formalmente os algoritmos utilizados. A partir do algoritmo de clusters são realizadas as análises de dinamismo e de capacidade da rede veicular.

Assim, as duas principais observações da análise realizada nesta tese consistem das seguintes características:

- *Revelações sobre o sistema público de transporte:* Ônibus têm rotas fixas, com trajeto em sua maior parte em ruas principais e avenidas importantes, enquanto táxis alcançam ruas secundárias, menos movimentadas. Os ônibus levam menos tempo do que os táxis para alcançar o próximo ponto de transferência. Dessa forma, combinar ônibus e táxis amplia a performance da rede, tanto em relação à capacidade quanto à área de cobertura;
- *Capacidade de transmissão:* É observada uma quantidade significativa de dados que podem ser descarregados através da rede veicular. Por exemplo, ao assumir uma banda de 1 Mbps para carga/descarga de dados dos veículos, o sistema pode transportar mais de 500 TB de dados diariamente.

1.2 Organização do Trabalho

O restante desta tese está organizado da seguinte forma. O Capítulo 2 apresenta o estado da arte e os trabalhos relacionados. O Capítulo 3 descreve a metodologia de coleta e os dados de posicionamento obtidos, além de iniciar a análise desses dados brutos. O Capítulo 4 mostra a metodologia utilizada para analisar os dados. No Capítulo 5 é realizada uma análise estrutural dos dados coletados, observando a

dinâmica da rede formada. O Capítulo 6 analisa as possibilidades de transferências de dados pelos veículos que circulam na cidade do Rio de Janeiro. Por fim, o Capítulo 7 conclui esta tese e apresenta os trabalhos futuros.

Capítulo 2

Estado da Arte

Neste capítulo são abordados os principais conceitos relacionados ao presente trabalho. Não foi encontrada na literatura uma análise de rede veicular composta por veículos de transporte públicos urbanos na cidade do Rio de Janeiro, porém, há trabalhos na literatura que utilizam redes veiculares em diversas cidades e contextos. Assim, neste capítulo são apresentados trabalhos da literatura relacionados ao transporte de dados por veículos em diversas situações e trabalhos onde são propostos sistemas que coletam, modelam e utilizam informações de deslocamento de frotas.

2.1 Transporte de dados por veículos

Um tipo específico redes móveis *ad hoc* (MANET – *mobile Ad Hoc Networks*) caracteriza-se quando a troca de dados ocorre entre veículos, conhecido como redes veiculares *ad hoc* (*Vehicular Ad Hoc Networks* – VANETs) [7], ou simplesmente redes veiculares. As redes veiculares podem utilizar padrões de comunicação especificamente definidos, utilizando a frequência definida de 5,9 GHz e o protocolo IEEE 802.11p [8], ou podem utilizar outros padrões de protocolos, como por exemplo o IEEE 802.11 de redes locais sem-fio e redes de telefonia celular.

Como não foram identificados trabalhos relacionados que contemplassem o assunto abordado na sua totalidade, o tema é dividido em quatro partes relacionadas ao transporte de dados por veículos. Inicialmente leva-se em consideração o transporte de dados, não importando o tipo de veículo ou a tecnologia de transmissão. Em seguida, são identificados os trabalhos que utilizam os mesmos tipos de veículos, porém que não analisam a capacidade de transmissão da rede ou que não utilizam todo o banco de dados de localização de veículos em suas análises. Posteriormente, são apresentadas as tecnologias de transmissão que são empregadas usualmente em transmissões de dados por veículos e a seção é concluída apresentando quais são os banco de dados de veículos disponíveis.

Assim, esta seção reúne trabalhos que abordam o transporte de dados por veículos e está subdividida em quatro subseções. A primeira subseção está relacionada a transmissões de dados por qualquer tipo de veículo, porém relacionados a dados de posicionamento dos mesmos. A segunda subseção contempla estudos de redes de transporte de veículos públicos urbanos, estudando aspectos como métricas relativas às viagens, protocolos de roteamento de mensagens e envio de informações. A terceira subseção está relacionada às tecnologias de transmissão sem-fio em redes veiculares. A quarta subseção descreve trabalhos relativos a bancos de dados de localização de veículos disponíveis na literatura.

Nenhum dos estudos a seguir contempla a utilização de múltiplas tecnologias de transmissão e o uso de uma base de dados com localização real de ônibus e táxis para determinar a capacidade de transmissão de uma rede sem-fio composta por esses veículos, contribuição realizada nesta tese.

2.1.1 Transmissões de dados por quaisquer veículos

Mitton *et al.* [9] analisam uma rede veicular que utiliza bicicletas para transportar dados entre estações de aluguel de bicicletas na cidade de Lyon, na França. Nesse trabalho, foi explorada a capacidade de transferência de dados por meio sem-fio. Foram utilizados dados reais de número de bicicletas alugadas, entretanto, foi utilizado um modelo para determinar a trajetória das bicicletas e não traços de mobilidade reais. Cada bicicleta pode transportar 12,49 GB diariamente e o conjunto de bicicletas chega a transportar 50 TB de dados diariamente.

Gorcitz *et al.* [10] modelam a rede de rodovias da França a partir de dados de quantidade de veículos que trafegam anualmente em segmentos das rodovias. Com esses dados, foram modeladas as viagens e modelada uma situação onde cada automóvel consegue transportar dispositivos de memória, contendo entre 250 GB e 1 TB, que devem ser descarregados ou substituídos no destino da viagem. Como resultado, foi possível determinar que em alguns trechos é possível transportar 1 PB de dados por 118 km em 8 horas. Ao equipar todos os veículos com armazenamento de 1 TB, é possível transferir 260 TB por hora, 200 vezes mais do que o método de transmissão por enlaces dedicados de Internet, o NetStitcher [11]. Dessa forma o transporte de dados é executado a partir de mídias, sem nenhuma interface de rede sem-fio para a troca de dados, sendo necessária a manipulação das mídias para a conclusão das transferências. Nesta tese, apesar do veículo também transportar fisicamente os dados, todas as transferências ocorrem através de meio sem-fio e os pontos de troca de dados são localizados no percurso dos veículos em pontos estratégicos.

Keranen *et al.* [12] propõem o uso de aviões para transportar dados. A partir

do mapa de voos mundial, foram observadas as viagens entre os maiores aeroportos do mundo. Nessa abordagem, cada aeronave é capaz de transportar até 450 GB. São estudadas as características das trocas de mensagens entre os aeroportos e a quantidade total de dados que podem trafegar entre alguns aeroportos. Partindo de Helsinque, na Finlândia, a banda para Los Angeles, nos Estados Unidos, é de 71 Mbps e chega a até 333 Mbps para Estocolmo, na Suécia. A modelagem é feita a partir de filas FIFO (*First In, First Out*), onde cada avião é responsável por transportar os dados e os aeroportos servem como infraestrutura para as trocas de dados entre os aviões. Este trabalho estuda protocolos de roteamento e trocas de mensagens e identifica a quantidade máxima de dados que poderiam ser distribuídas entre os aeroportos ao considerar uma conexão por fio com os aviões em seu tempo no solo. Além disso as viagens de avião ocorrem de maneira previsível, tendo um trajeto predeterminado e somente um horário de decolagem e um horário de aterrissagem. O presente trabalho analisa dados de ônibus e táxis, que não tem a mesma previsibilidade em seu deslocamento e utiliza a transferência de dados por meio sem-fio.

Zarafshan-Araki *et al.* [13] propõem o uso de trens para transportar dados entre diversas estações. O trabalho se concentra em criar um algoritmo para determinar a quantidade de dados que podem ser transportados entre as estações, sem que o armazenamento nos trens seja extrapolado. Cada trem é equipado com 1 TB de capacidade, formado por 4 discos de 250 GB. Os dados são transferidos quando o trem chega nas estações através da troca física dos discos. A partir de dados reais de horários de trens, a simulação evidencia bandas de até 1,4 Tbps entre determinadas estações.

Santos *et al.* [14] propõem o transporte de dados na bacia amazônica utilizando barcos e redes sem-fio. Cada barco é equipado com um ponto de acesso de rede sem-fio utilizando a tecnologia IEEE 802.11g. A partir de dados reais, coletados entre os barcos e de dados do transporte fluvial na bacia amazônica, foi modelada a capacidade de transporte de dados. Ao trocar dados somente quando ocorrem os encontros entre os barcos, é possível transferir até 93 GB de dados utilizando a rede IEEE 802.11g. Nesta tese, de forma similar, é analisado o transporte de dados utilizando redes sem-fio, porém não é definida uma tecnologia específica para a transferência de dados, deixando assim como variáveis de entrada o alcance de rádio e banda de transferência.

Os artigos [9, 10, 12–14] propõem métodos distintos para transportar dados através de diversos tipos de veículos. Cada um deles conclui que é possível transportar da ordem de terabytes de dados em cada um dos seus cenários, de forma similar ao presente trabalho. Entretanto, nesta tese utiliza-se um banco de dados de veículos maior e escolheu-se focar na identificação de regiões para comunicação

com os veículos, contabilizando o tempo que estes permanecem em contato com um ponto de acesso. Essa forma de análise é distinta das anteriores pelas características dos veículos, pela quantidade de veículos e pelo meio de transmissão de dados utilizado. Essas características levam a um método de análise próprio, sendo necessário o uso de ferramentas para análises de grandes dados, não sendo possível determinar um modelo único para caracterizar o transporte dos dados.

2.1.2 Veículos de transporte público urbanos

Mazloumi *et al.* [15] tratam da variação temporal no tempo de viagens de ônibus de acordo com o horário do dia na cidade de Melbourne, na Austrália. Foi observado que o tempo de viagem tende a uma distribuição normal em horários movimentados, com muitas partidas de ônibus. Entretanto, para viagens em períodos fora do horário de pico, observa-se um comportamento lognormal para o tempo de viagem. O número de sinais de trânsito tem um impacto maior no tempo de viagem de um ônibus do que o número de paradas do mesmo. O artigo se limita a observar as variações de tempo, não explorando como o deslocamento dos ônibus pode ser utilizado para outros fins.

Sroush *et al.* [16] montam um *testbed* composto de 40 ônibus em um campus que foram equipados com diversas interfaces de comunicação sem-fio, incluindo GPS, modem 3G e rádio IEEE 802.11g e pontos de acesso previamente distribuídos no ambiente. Os pontos de acesso distribuídos são conectados à Internet e não tiveram as suas posições devidamente otimizadas. O artigo mostra como foram coletados os dados e faz algumas análises dos dados coletados, que foram a banda de transferência dos rádios IEEE 802.11g e celular e a duração do contato entre os ônibus.

Zhou *et al.* [17] utilizam um aplicativo de celular e dados de localizações dos veículos para prever a demanda de ônibus na cidade de Yantai, na China. Os dados obtidos mostram um maior uso dos transportes públicos de passageiros nos horários de entrada e saída do trabalho, além de diferenças no padrão de usos em dias com ou sem chuvas e em dias úteis ou finais de semana. Os dados ainda mostram que o tempo de embarque dos passageiros nos ônibus é maior quando todos os assentos estão ocupados em relação a quando ainda existem lugares para sentar. É utilizado um banco de dados de 22 semanas com 1326 ônibus, com atualizações de posição a cada 2 minutos e contendo informações de lotação do coletivo. A modelagem da quantidade de passageiros no coletivo é efetuada através do tempo parado nos pontos de ônibus e da informação de lotação do veículo. Um algoritmo para prever a demanda dos ônibus é criado, entretanto esse algoritmo é treinado especificamente para a cidade de Yantai e para o padrão de deslocamento das pessoas daquela cidade. O trabalho explora a quantidade de pessoas que utilizam os ônibus, porém não está

relacionado com a utilização desses ônibus para transportar dados.

Sede *et al.* [18] propõem um protocolo de roteamento capaz de utilizar informações das linhas de ônibus como parâmetro de roteamento. Para a simulação, foram utilizados 700 ônibus de 30 linhas. O método é utilizado para rotear dados entre as linhas de ônibus, desta forma, agrupando as linhas para diminuir a complexidade da análise. Comparado a protocolos de disseminação epidêmica e de primeiro contato, o algoritmo é capaz de ganhos expressivos na transmissão de informações de roteamento ao utilizar tais informações de linhas. Somente o desempenho do protocolo é examinado, não sendo analisada a quantidade de dados transferidos pelos ônibus. O algoritmo é utilizado para troca de dados entre veículos, não levando em consideração locais para troca de informações. Dessa forma, o algoritmo somente pode ser utilizado se for adaptado para considerar esses locais de troca de informações. O algoritmo apresenta uma desvantagem por ser mais complexo do que o necessário para o roteamento na rede proposta, como mostra a análise realizada na Seção 5.2. Além disso, as informações de linha de ônibus dos dados coletados não são confiáveis, onde uma considerável parcela dos veículos não informa corretamente qual é a sua linha de ônibus.

Doering *et al.* [19–21] estudam os padrões de mobilidade dos ônibus, principalmente as características de contato entre eles, como a quantidade de ônibus ativos, velocidade, taxa de atualização, duração de contato entre veículos. Além disso, é analisada a duração do contato em relação ao tipo de deslocamento, onde o deslocamento pode ser de 5 tipos: parado, perseguindo, encontro, interseção e não identificado. Cada um dos padrões de deslocamento tem uma característica de tempo de contato, com variações grandes de tempo médio de contato ao utilizar rádios de baixo alcance (50 m) e com pequenas variações ao utilizar rádios de longo alcance (500 m). Os dados analisados contem a localização de 1600 ônibus das cidades de Seattle e de Chicago, ambas nos Estados Unidos, com atualizações entre 20 e 40 segundos, sendo interpolados para obter a resolução de 1 segundo de intervalo de atualização. Os autores propõem um algoritmo para roteamento de pacotes a partir das análises dos tempos de contatos dos ônibus e da direção de encontro. Nenhuma análise da capacidade da rede formada pelos ônibus foi realizada.

Em um trabalho posterior Doering *et al.* [22] usam os mesmos dados de deslocamento de ônibus para obter métricas específicas de cada uma das cidades, como número de ônibus ativos, velocidade, distância entre localizações recebidas, densidade de ônibus, e comparar seus sistemas de transporte. Uma análise semelhante baseada nesse trabalho é realizada para os dados de transporte coletados na cidade do Rio de Janeiro no Capítulo 3 porém são comparados táxis e ônibus e não de duas cidades distintas. Além de informar as características das cidades com relação a essas métricas, é indicado o uso de redes tolerantes a atraso para a coleta de in-

formação espalhadas pela cidade e que o poder computacional necessário para as análises e simulações dos encontros entre veículos dessas redes tem uma magnitude elevada a ponto de precisar de novas abordagens para a manipulação do problema.

Jiang *et al.* [23] analisam o transporte de informações entre veículos e pontos de infraestrutura ou *roadside units* (RSU). Os veículos são divididos em duas categorias, carros e ônibus. Os carros somente possuem interface de rádio para comunicação com outros veículos e com as RSUs, porém os ônibus tem essa interface e uma segunda interface de longo alcance para contato com as RSUs, dessa forma criando uma espinha dorsal ou *backbone* para as transmissões de dados. Ao utilizar a capacidade de transmissão dos veículos com os elementos da via, os autores criam um mecanismo de roteamento capaz de determinar o trajeto de um veículo e a possibilidade de transferências de dados. O sistema projetado precisa da localização de tempo real de todos os veículos, disseminada pela rede, para ser capaz de estabelecer as rotas desejadas e do atual percurso dos veículos. Como resultado, o algoritmo é capaz de entregar em poucos segundos os dados requisitados pelos usuários. A simulação é executada a partir de dados sintéticos e com uma quantidade baixa de veículos (100) e de tempo (5 minutos) e nenhuma análise de capacidade de transmissão de dados é executada, somente características do protocolo de roteamento são desenvolvidas. O protocolo de roteamento desenvolvido é excessivamente complexo, necessitando do conhecimento exato das localizações de todos os veículos, o tornando incompatível como protocolo de roteamento para o presente trabalho.

Naboulsi *et al.* [24] usam dados de localização de veículos sintetizados através de modelos da cidade de Colônia, na Alemanha. A partir desses dados, foram feitas análises de redes complexas, buscando localizar no grafo os maiores componentes conectados existentes. Foi observado que o número de componentes conectados gigantes é baixo e aparece somente em horários de intenso movimento, como nos horários da manhã. Outra observação realizada é que o número de veículos influencia diretamente a quantidade de componentes conectados e não o horário, ou seja, se horários distintos tem a mesma quantidade de veículos em circulação, as suas características são semelhantes. Por último, o número de componentes e o tamanho dos mesmos varia de acordo com o raio de alcance da comunicação, onde um alcance maior aumenta o tamanho dos componentes conectados e diminui o número total deles. Nesse trabalho não são analisadas as características dos dados gerados nem a capacidade de transporte de dados entre os veículos, porém mostra que a análise da rede complexa é capaz de identificar padrões de deslocamento e características dos contatos entre os veículos que não são visualizadas facilmente por outros meios de análise.

Os artigos [15–24] analisam redes veiculares formadas por ônibus ou outros veículos, se concentrando em identificar padrões na rede ou propor algoritmos para

a transferência de mensagens ou algoritmos de roteamento. São realizados estudos sobre os protocolos de roteamento para diversos cenários, porém não é estudada a capacidade de transferência de dados das redes formadas por veículos e pontos de acesso, indicando a localização dos pontos de acesso nos locais de maior capacidade de transferência de dados.

2.1.3 Tecnologias de transmissão sem-fio

Araniti *et al.* [25] apresentam uma extensa pesquisa sobre comunicações de veículos, comparando a tecnologia de redes celulares LTE (*Long Term Evolution*) a outras tecnologias, notadamente aos padrões de redes locais sem-fio IEEE 802.11g e a IEEE 802.11p. Essa pesquisa foca nas aplicações de segurança, como avisos e informações sobre as condições da estrada. O trabalho indica que o LTE pode ser utilizado em redes veiculares, mesmo que a latência seja maior, devido à elevada distância entre os equipamentos sem-fio em comparação a outras tecnologias. Além disso o LTE tem a vantagem de poder ser usado em combinação com, por exemplo, o IEEE 802.11p.

A comunicação de dispositivo a dispositivo (D2D – *device-to-device*) foi acrescentada posteriormente no padrão LTE. Piro *et al.* [26] analisam a utilização do LTE-D2D em redes veiculares com foco na troca de mensagens de controle (*Cooperative Awareness Messages – CAM*) entre veículos. Nesse trabalho é investigado o número de veículos simultâneos e o alcance que a tecnologia é capaz de suportar. É identificado que a rede suporta a quantidade de comunicação necessária entre veículos, com até 25 veículos por metro nas condições simuladas. Os enlaces D2D podem ser formados entre distâncias necessárias para a transmissão de informações de cooperação entre veículos.

Campolo *et al.* [5] apresentam uma visão geral da possível evolução das redes veiculares sem-fio, desde a atual rede IEEE 802.11p até o 5G e protocolos futuros de comunicação. O estudo define fases de implementação. Na primeira fase, somente avisos são enviados, como condições de estrada e veículos de emergência próximos. Esta fase é compatível com a tecnologia atual do IEEE 802.11p. As últimas fases, baseadas na evolução tecnologias atuais e nas novas tecnologias como LTE-A, D2D, 5G, serão capazes de suportar o uso de comboio e até a condução autônoma dos veículos. Essa última fase também abre a possibilidade para outros usuários, como os pedestres que possuem um smartphone, de participar da rede. Esta última abordagem conectará veículos e “coisas” (por exemplo, sensores, câmeras), abrindo uma nova relação entre dispositivos.

Os trabalhos relacionados nesta seção ([5, 25, 26]) mostram a viabilidade de usar diversas tecnologias de redes sem-fio em cenários de redes veiculares e a necessidade

de buscar novas formas de comunicação, principalmente para aliviar as tecnologias de transmissão a serem empregadas em aplicações de segurança. Estes trabalhos fornecem uma visão do possível futuro das transmissões de dados nas redes veiculares, além de confirmar a importância dos dispositivos conectados no ambiente urbano.

Os trabalhos até agora apresentados utilizaram algum tipo de dado de localizações dos veículos. Na próxima seção são apresentados os tipos possíveis de dados, sintéticos ou reais, e os principais bancos de dados disponíveis.

2.1.4 Bancos de dados de deslocamento disponíveis

Para a análise de uma rede veicular, onde os participantes são capazes de se deslocar pelas vias, é imprescindível possuir a localização dos veículos que a compõe. Estão presentes na literatura dois tipos de dados de traços de localização e movimentação de veículos: dados gerados através de modelos matemáticos de deslocamento e dados de trajetórias coletados do mundo real.

Para os dados sintéticos, normalmente são escolhidas regiões, como uma cidade ou um bairro, em que todas as ruas e sinalizações de trânsito foram mapeadas. A partir de modelos matemáticos de deslocamento são gerados traços de deslocamento dos veículos em um intervalo de tempo. Em muitos casos, a escolha dos parâmetros – quantidade de carros, velocidade, rotas, entre outros – e do modelo de deslocamento não é realizada de forma ótima pois são atributos de alta complexidade e que apresentam grande variação em relação ao tempo, ambiente, entre outros. Isso não permite obter o rigor estatístico necessário para representar a distribuição em larga escala do tráfego, mas apenas captura com confiança curtos espaços de tempo ou pequenas áreas urbanas. Por essas razões, os dados disponibilizados com essa abordagem geralmente são limitados a poucos quilômetros de área e a curtos intervalos de tempo [27, 28].

Os dados coletados no mundo real são obtidos a partir da gravação direta das posições de veículos durante o seu deslocamento. Dados desse tipo normalmente são obtidos pelo monitoramento de frotas de empresas [29]. Não é possível a coleta de dados de todos os veículos que circulam em uma cidade por vários motivos, inclusive pelo direito à privacidade dos motoristas e pela heterogeneidade dos veículos: nem todos possuem equipamentos de GPS ou a tecnologia de comunicação e armazenamento necessária para o compartilhamento dos dados de deslocamento.

Na literatura estão disponíveis dados de deslocamento de frotas de ônibus, de táxis e de carros de empresas. Uma das primeiras fontes de dados coletados através do sistema de localização dos veículos a ser disponibilizada publicamente foi a dos ônibus de Seattle [30]. Este traço apresenta o movimento de 1.200 ônibus, com um intervalo de coleta de dois minutos durante duas semanas. Dados similares

foram colhidos em Chicago, com 1.648 ônibus e com intervalos de atualizações de quarenta segundos durante 18 dias [20]. Outra fonte de dados semelhante foi obtida sobre o testbed DieselNet [31], que consistiu em equipar uma pequena quantidade de veículos (trinta) com equipamentos para monitorar o seu deslocamento a cada segundo por 60 dias. Outras análises, com táxis e frotas empresariais, foram feitas em São Francisco [32] e em Xangai [18]. Uma análise mais detalhada dos dados de deslocamento disponíveis na literatura é feita por Uppoor *et al.* [29].

Nesta tese, especificamente, são utilizados dados reais de deslocamento de ônibus e de táxis da cidade do Rio de Janeiro para determinar as possibilidades de transmissão de dados, com duração de 1 semana para os dados de localização dos ônibus e de 1 dia para os dados dos táxis. Por isso, é importante estudar o estado da arte para os estudos relacionados a redes veiculares formadas por ônibus e táxis em outras cidades [33].

2.2 Aplicações

Diversas aplicações podem ser executadas sobre uma rede veicular. Assim, um dos objetivos desta tese é analisar a capacidade da rede independente da tecnologia de transmissão, para que, a partir do resultado da análise, possa ser descoberta a viabilidade de uma determinada aplicação. Dessa forma, não é definida uma aplicação específica para a rede. No entanto, esta seção traz algumas das possibilidades presentes na literatura de aplicações que podem ser aproveitadas nesse canal alternativo às redes celulares para o transporte de informações.

2.2.1 Cidades inteligentes

O tema cidades inteligentes traz um novo paradigma para o desenvolvimento sustentável das cidades. Como definido por Albino *et al.* [3], uma cidade é considerada inteligente quando contém determinadas características, como uma estrutura de rede que permita políticas eficientes e o desenvolvimento social e cultural, ênfase no desenvolvimento urbano empresarial e atividades criativas para promover o crescimento urbano, inclusão da comunidade e dos residentes no desenvolvimento urbano e um ambiente natural como componente estratégico para o futuro.

Essas características apontam para um futuro mais colaborativo para as cidades inteligentes, no qual todos os cidadãos estão ligados e suas ações e costumes são levados em consideração para o desenvolvimento da cidade. Dessa forma, destacam-se atualmente alguns serviços inteligentes com essas características. Albino *et al.* enumeram alguns serviços inteligentes relacionados ao uso de Internet ou à transmissão de dados entre entidades, dentre os quais se destacam:

- redes elétricas inteligente (*smart grids*) para a coleta e tarifa diferenciada para o consumo de gás e eletricidade;
- sistemas inteligentes de drenagem e captação de dejetos;
- programa de monitoramento das condições climáticas;
- serviço de Internet Wi-Fi para a população;
- controle on-line dos sistemas de transporte;
- monitoramento remoto de pacientes, como por exemplo incluindo dados da condição do coração.

Como a cidade, as relações sociais e a tecnologia estão em constante evolução, esses serviços providos nas cidades inteligentes podem evoluir com o tempo. Como exemplo, podemos destacar que novas tecnologias possibilitarão que novas variáveis sejam monitoradas, trazendo um melhor conhecimento da cidade e melhorando as relações sociais.

Sistemas de comunicação têm um papel fundamental nas cidades inteligentes, possibilitando prover os serviços necessários de acordo com a demanda. Porém, a realização dessas tarefas depende da disponibilidade de uma infraestrutura de comunicação confiável e onipresente, com capacidade e tempo de resposta adequados às aplicações.

Diversas aplicações para o transporte de dados através de veículos de transporte público, como ônibus e táxis, podem ser imaginadas sem entretanto haver uma aplicação chave para esse sistema. A seguir são exploradas algumas áreas e aplicações que podem se beneficiar do transporte de dados proposto.

2.2.2 Distribuição de informações para veículos

Jia *et al.* [34] utilizam as redes veiculares formadas por ônibus para exibir anúncios internamente para os passageiros e externamente para carros particulares ao redor dos ônibus. Os anúncios são exibidos ou enviados aos carros quando os ônibus se aproximam dos veículos, dessa forma, o ônibus é responsável pela transmissão de dados para os carros particulares ao redor. A rede formada entre os ônibus é utilizada para a disseminação dos anúncios, atualizando-os de acordo com a demanda. O artigo foca na modelagem da forma de exibição e capitalização dos anúncios através da contagem da disseminação das transmissões de ônibus para os carros. A partir de uma modelagem e de dados genéricos de quantidade de veículos por via, são exibidos resultados de exibição de anúncios. O trabalho não apresenta a topologia de rede utilizada para o transporte nem executa uma análise dos dados transmitidos para a exibição dos anúncios.

Schildt *et al.* [35] propõem estratégias para remunerar os usuários que fazem parte de uma rede de transmissão de dados, como uma DTN. O sistema cria incentivos financeiros através de anúncios oferecendo descontos em produtos ou mesmo produtos grátis. O artigo mostra cenários onde essas recompensas podem ser utilizadas, incluindo a transmissão de dados pelos usuários do sistema. Entretanto, o objetivo do artigo é mostrar a viabilidade de disseminar anúncios para usuários através de redes sem-fio e dispositivos móveis.

Os múltiplos tipos de conexões e recursos computacionais embarcados nos veículos também motivaram pesquisas recentes na área de virtualização. Como exemplo, Li *et al.* em [36] exploram os conceitos de redes sociais aplicadas às redes veiculares. Os autores consideram redes veiculares definidas por software, onde os veículos podem se juntar a grupos para executar funções específicas, como alertas em estrada ou formação de comboios. Já Baron *et al.* [37] exploram a virtualização dos recursos de computação presentes nos veículos e a migração de serviços entre veículos próximos usando comunicações V2V (*vehicle-to-vehicle*).

Mezghani *et al.* [38] criam um método de distribuição de informações em uma rede veicular a partir do conceito de redes sociais. Dados de gerenciamento de tráfego ou mesmo dados produzidos por aplicações de entretenimento para os passageiros são transportados por veículos. Ao utilizar os conceitos de redes sociais, como preferências dos usuários, estimativa de relevância do conteúdo e caches inteligentes, os dados gerados e requisitados podem ser melhor distribuídos entre os participantes, reduzindo o tempo de espera para receber o dado.

2.2.3 Métodos de transporte de grandes dados

Os métodos de transmissão de grandes dados estão voltados para o transporte de grande quantidade de dados entre centros de dados, entre consumidores ou somente abordando novos métodos de comunicação. A abordagem mais comum é descobrir maneiras de otimizar as transferências por meio de agendamentos inteligentes utilizando a ociosidade de enlaces, como a solução NetStitcher [11] e a StorkCloud [39].

Outra alternativa existente na literatura para o transporte de grandes dados é seu envio através de meios físicos, como discos rígidos ou mídias ópticas (por exemplo, DVD) [40]. Essas abordagens são baseadas no transporte de discos rígidos por serviços postais, o que muitas vezes torna esse transporte de dados mais caro do que o envio por enlaces de comunicação [41]. Dessa forma, o transporte de informações deve ser feito de uma maneira sustentável economicamente, não sendo custoso para quem quer transportar os dados e ainda deve ser rápido o suficiente para conseguir atender a demanda.

Neste capítulo foram explicitadas propostas que aproveitam os meios de trans-

portes existentes para construir redes veiculares para o transporte de dados. Diferentes tipos de veículos, como bicicletas [9], automóveis [10], aviões [12], ônibus [16], trens [13] ou ainda barcos [14] são estudados como nós móveis (mulas) de transporte de dados. Embora cada tipo de veículo e rede veicular possua suas peculiaridades, todos os trabalhos mostram que é possível e viável o transporte físico dos dados, observadas as suas limitações [42–44].

No Capítulo 3 serão descritos os dados coletados dos ônibus e os dados dos táxis disponibilizados pela empresa 99. A partir de todos esses dados, será observado o comportamento do transporte público da cidade do Rio de Janeiro.

Capítulo 3

Coleta dos Dados de Veículos Públicos Coletivos da Cidade do Rio de Janeiro

A análise dos dados de localizações dos veículos é essencial para se identificar as características dos deslocamentos dos veículos. Ao observar as métricas como quantidade de veículo, localização dos veículos e horário de atividade é possível indicar padrões presentes na movimentação os veículos. Neste capítulo, na Seção 3.1, são mostrados dados de interesse da cidade do Rio de Janeiro e dos veículos analisados. Na Seção 3.2, são analisados os dados brutos de localização dos ônibus e as suas características, fazendo uma análise geral dos dados na Seção 3.2.4, onde os padrões presentes nos dados coletados são evidenciados. Na Seção 3.3, é feita uma análise semelhante dos padrões de deslocamento porém a partir dos dados de posicionamento dos táxis.

3.1 Infraestrutura de transporte na cidade do Rio de Janeiro

Apesar de existirem bancos de dados que capturam movimentos de veículos em diversas cidades do mundo, as cidades possuem características distintas em termos de malha rodoviária e de sistema de transportes. A cidade do Rio de Janeiro é uma das metrópoles do Brasil. Possuía, em 2015, aproximadamente 6,4 milhões de habitantes [45]. O Rio de Janeiro tem aproximadamente 1.200 km² de área e mais de 11.000 km de vias [46], por onde circularam, em 2014, 2.576.947 veículos, sendo destes 17.723 ônibus e 32.000 táxis [47].

De acordo com dados operacionais sobre o transporte coletivos urbanos, obtidos na página da Federação das Empresas de Transportes de Passageiros do Estado do

Rio de Janeiro [48], em outubro de 2014 havia 725 linhas de ônibus ativas com uma frota de 9.028 veículos (ônibus). Naquele mês foram feitas 1.832.461 viagens, nas quais os ônibus percorreram mais de 72 milhões de quilômetros e transportaram mais de 117 milhões de passageiros. A Tabela 3.1 mostra os dados mês a mês para o ano de 2014 [6].

Tabela 3.1: Dados operacionais mensais da cidade do Rio de Janeiro (2014).

Mês	Linhas	Frota	Viagens	Quilômetros percorridos	Total de passageiros	Litros de diesel
jan/14	699	8.761	1.511.333	65.037.199	100.036.919	22.656.166
fev/14	693	8.713	1.358.479	58.588.166	99.300.618	20.542.490
mar/14	707	8.790	1.420.360	60.917.036	101.168.869	20.938.619
abr/14	713	8.830	1.413.597	60.714.477	101.214.431	21.519.224
mai/14	712	8.884	1.353.768	57.999.247	98.156.418	20.697.478
jun/14	707	8.917	1.416.578	59.908.800	96.414.412	20.635.768
jul/14	729	8.994	1.627.717	66.397.102	106.564.705	21.933.311
ago/14	731	9.022	1.637.631	65.820.418	113.245.928	22.293.446
set/14	727	8.947	1.629.803	64.298.419	113.737.253	21.793.742
out/14	725	9.028	1.832.461	72.943.133	117.104.451	24.312.000
nov/14	721	9.046	1.594.881	62.159.160	107.592.882	20.829.126
dez/14	724	9.065	1.695.289	64.296.747	109.378.269	24.397.603

Além disso, na cidade do Rio de Janeiro atualmente circulam 32 mil táxis, de acordo com dados de abril de 2016 [47]. O aplicativo móvel 99 forneceu, para o presente trabalho, dados de localização dos táxis nele cadastrados. O 99 é o maior aplicativo para solicitar táxis pelo celular no Brasil. Fundada em 2012, a empresa tem a maior frota de taxistas cadastrados e está presente em mais de 300 cidades [49]. Os dados cedidos pela 99 referem-se ao dia 07/10/2014 e apresentam 24 horas de posicionamento de 5.803 táxis, o que representa 18% da frota de veículos da cidade do Rio de Janeiro.

3.2 Dados dos ônibus urbanos da cidade do Rio de Janeiro

Dados em tempo real das posições informadas pelos ônibus, atualizados a cada minuto, além de dados de pontos de ônibus, trajetos de ônibus e outros dados de mobilidade da cidade são disponibilizados pela prefeitura do Rio de Janeiro no portal de dados abertos da cidade [50]. Nesse portal são disponibilizadas informações sobre a mobilidade dos ônibus, incluindo os pontos de paradas [51], os pontos geográficos dos trajetos [52] e os dados de localização em tempo real [53].

O autor dos dados é a IplanRio, que segundo o site [54], é a Empresa Municipal

de Informática da Cidade do Rio de Janeiro. O mantenedor da base de dados é a Fetranspor e a frequência de atualização é a cada minuto [53].

O guia de utilização do banco de dados [55] informa os campos para a leitura exibidos na Tabela 3.2. Os dados são compostos pelo horário de coleta, pela identificação do ônibus, pela linha à qual o ônibus está designado, por informações de localização (latitude e longitude) e pela velocidade.

Tabela 3.2: Metadados associados ao dataset – GPS ônibus.

Campo	Descrição	Tipo	Tamanho
DataHora	Data e hora da coleta do dado	DATETIME	23
Ordem	Identificação alfanumérica encontrada na lateral dos ônibus	VARCHAR	7
Linha	Linha do ônibus	VARCHAR	7
Latitude	Latitude do ônibus na coleta (GPS, WGS84)	NUMERIC	11
Longitude	Longitude do ônibus na coleta (GPS, WGS84)	NUMERIC	11
Velocidade	Velocidade do ônibus na hora do coleta do dado	NUMERIC	6

Esse conjunto de dados encontra-se disponível para download no formato CSV, a partir do acesso ao site da Internet: <http://dadosabertos.rio.rj.gov.br/apiTransporte/apresentacao/csv/onibus.cfm>.

O Código 3.1 mostra uma porção do resultado de uma consulta ao banco de dados da prefeitura realizada às 16:54h do dia 29 de fevereiro de 2016. A consulta informa 6.679 linhas, porém, por simplificação, foi somente exibido o início e o final da consulta. Em cada consulta é apresentada a última posição de todos os ônibus que enviaram pelo menos uma localização naquele dia. Como um ônibus em atividade envia uma atualização por minuto, é esperado que o campo “DATAHORA” seja sempre atualizado e como um horário próximo ao horário da consulta, porém é observado que nem sempre isso acontece. Quando um ônibus não envia mais os dados de localização, o seu último tempo e posição conhecidos são exibidos a cada consulta realizada, como apresentada nas 10 entradas do Código 3.1. Dados desse tipo são recebidos em todas as consultas subsequentes, sendo considerados dados duplicados de localização.

Código 3.1: Parte dos Dados recebidos em uma consulta

```
{ "COLUMNS" :
[ "DATAHORA" , "ORDEM" , "LINHA" , "LATITUDE" , "LONGITUDE" , "VELOCIDADE" ] ,
"DATA" : [
["02-29-2016 00:00:43" , "A63536" , "" , -22.867781 , -43.258411 , 0.0] ,
["02-29-2016 00:22:41" , "B58112" , 313.0 , -22.83814 , -43.284729 , 55.0] ,
["02-29-2016 01:08:49" , "B75657" , "" , -22.83217 , -43.324539 , 0.0] ,
["02-29-2016 01:42:30" , "B10151" , "" , -22.81476 , -43.187351 , 0.56] ,
["02-29-2016 02:08:50" , "B63025" , 942.0 , -22.79229 , -43.293819 , 0.0] ,
["02-29-2016 02:49:53" , "A48038" , 106.0 , -22.910654 , -43.270138 , 0.0] ,
```

```

["02-29-2016 10:21:55", "A48105", 415.0, -22.984398, -43.202934, 0.0],
["02-29-2016 10:21:59", "A48090", "", -22.910608, -43.207119, 7.59],
["02-29-2016 15:15:26", "D86242", 884.0, -22.98402, -43.694817, 29.63],
["02-29-2016 15:15:26", "D86245", 864.0, -22.902296, -43.555393, 5.93],
...
["02-29-2016 16:53:45", "D87791", 2336.0, -22.84148, -43.371109, 0.0]
["02-29-2016 16:53:46", "A27556", 439.0, -22.919531, -43.22142, 15.0],
["02-29-2016 16:53:46", "A27562", "", -22.91383, -43.236301, 0.0],
["02-29-2016 16:53:46", "A27589", 432.0, -22.98344, -43.228191, 0.0],
["02-29-2016 16:53:46", "A27641", 548.0, -22.99371, -43.256088, 3.0],
["02-29-2016 16:53:46", "A27681", 432.0, -22.91432, -43.19717, 0.0],
["02-29-2016 16:53:46", "A29007", 473.0, -22.895639, -43.219082, 4.0],
["02-29-2016 16:53:46", "A29044", 472.0, -22.952221, -43.181702, 29.0],
["02-29-2016 17:04:03", "B25525", "", -22.90217, -43.29847, 0.0]]
]}

```

3.2.1 Coleta de dados de localização dos ônibus

Neste trabalho, foram coletados dados de GPS dos ônibus da cidade do Rio de Janeiro entre os dias 23 de setembro de 2014 e 01 de dezembro de 2014. Nesse intervalo, alguns dias tiveram a coleta de dados prejudicada. Entre os dias 24 e 28 de outubro de 2014 e os dias 12 e 19 de novembro não foi possível acessar o banco de dados, portanto, não existe a informação de localização desses períodos. Os dados coletados no intervalo somam mais de 13 GB, com um tamanho médio de 300 MB por dia.

As consultas foram realizadas a cada minuto, armazenando os dados não duplicados encontrados em cada dia em um arquivo distinto. Para as análises realizadas nessa tese foi escolhido o dia 07 de outubro de 2014, uma terça-feira, por ser o dia com mais dados de localização presentes no arquivo. Ou seja, não houve interrupções em nenhum momento na coleta dos dados nessa data. Para as análises que correspondem ao período de uma semana, foi escolhido o período de 06 de outubro, uma segunda-feira, até 12 de outubro de 2014, um domingo. O dia escolhido tem 6.226.714 entradas de dados, com o tamanho de 346 MB, e a semana escolhida tem 34.978.907 entradas de dados, com o tamanho de 1,9 GB. Foram coletados dados de 12.456 ônibus distintos.

A Figura 3.1 mostra todos os dados coletados no dia 07 de outubro de 2014, onde cada ponto representa um dado de posicionamento de um ônibus recebido nesse dia. É possível identificar visualmente que os ônibus tem a tendência de trafegar em vias importantes da cidade, que ligam as diversas regiões. Também é possível observar a concentração elevada de pontos no centro do canto direito da figura, o que representa

o centro comercial da cidade, e que os pontos estão espalhados por toda a cidade, menos nas regiões de floresta.

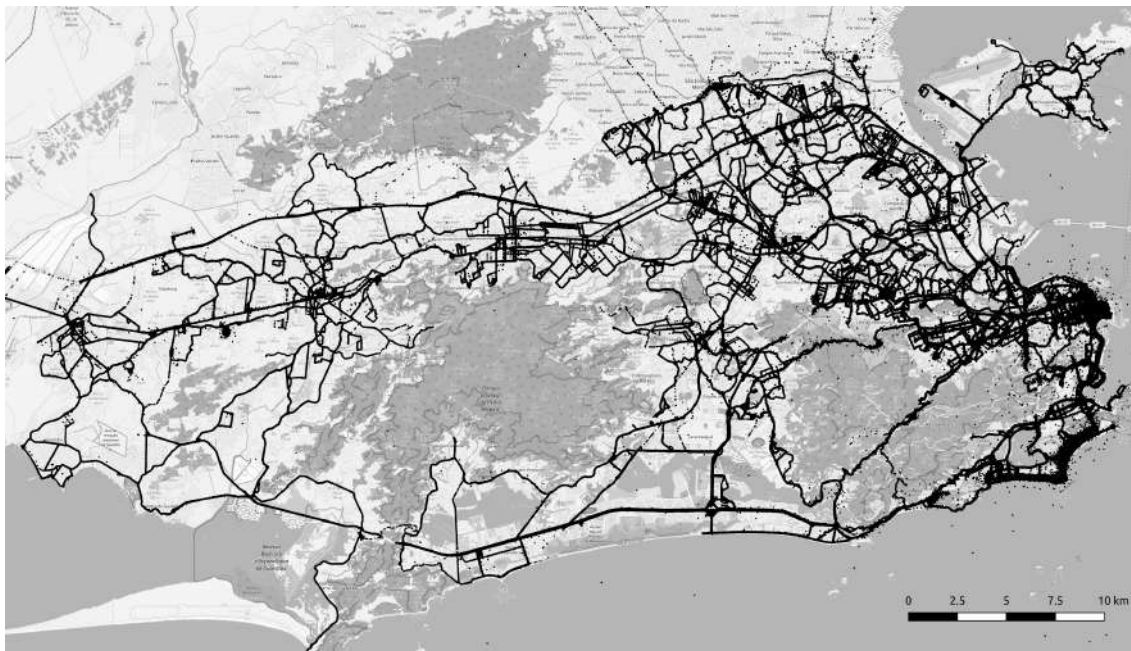


Figura 3.1: Mapa com a localização dos ônibus no dia 07/10/2014.

3.2.2 Dados de passageiros

A prefeitura do Rio de Janeiro disponibiliza algumas estatísticas mensais sobre as linhas de ônibus [56]. Como exemplo, há dados de quantidade de quilômetros rodados, de quantidade de passagens pagas e de gratuidades, além do total de passageiros transportados por cada uma das linhas.

A Tabela 3.3 mostra a quantidade mensal de passageiros, no mês de outubro de 2014, que circularam nos ônibus. Por questão de simplicidade, somente 10 linhas aleatórias são exibidas. A quantidade total de quilômetros percorridos em outubro de 2014 é equivalente a 6.000 vezes a malha viária da cidade, transportando 16 vezes o número de habitantes da cidade do Rio de Janeiro por mês.

3.2.3 Identificação dos erros de localização

O banco de dados possui características singulares em consequência da forma como os dados são coletados. Como os ônibus enviam os dados através da rede celular, os dados de localização não são sincronizados, ou seja, não têm o mesmo horário de coleta. Além disso, podem também não ser enviados em sua totalidade, o que ocorre quando, por exemplo, um ônibus não consegue contato com a rede de telefonia móvel.

Tabela 3.3: Quantidade de viagens mensal - Outubro de 2014.

Linha	Qtd km	Viagens Pagas	Viagens Gratuitas	Total de Passageiros
371 -PRAÇA DA REPÚBLICA X PRAÇA SECA	414.267	697.874	212.492	910.366
363 -PRAÇA XV X VILA VALQUEIRE	320.064	453.838	128.105	581.943
415 -USINA X LEBLON (CIRCULAR)	338.949	610.077	111.397	721.474
624 -MARIÓPOLIS X PRAÇA DA BANDEIRA	346.746	511.923	183.195	695.118
433 -VILA ISABEL X GAVEA (VIA COPACABANA)	258.543	408.145	89.042	497.187
550 -CIDADE DE DEUS X GÁVEA	307.546	575.605	122.315	697.920
2336 -CAMPO GRANDE X CASTELO	500.978	150.380	110	150.490
474 -JACARÉ X JARDIM DE ALAH	371.158	439.339	92.095	531.434
353 -RODOVIÁRIA X CIDADE DE DEUS	235.377	380.970	114.306	495.276
886 -FREGUESIA X BARRA DA TIJUCA	68.978	71.164	15.045	86.209
Total - considerando todas as linhas	65.576.563	80.388.644	19.334.869	99.723.513

A partir dos dados de localização dos ônibus plotados na Figura 3.1 é possível identificar visualmente pontos que estão localizados em locais indevidos, como no oceano ou no meio de uma floresta. Com o intuito de verificar alguns erros de GPS identificados, foi realizada uma análise manual de sete locais com erros. As Tabelas 3.4 a 3.10 detalham as localizações encontradas.

A Tabela 3.4 mostra a localização do ônibus de ordem D867990, que apresenta 4 localizações seguidas no mesmo lugar do oceano. Anteriormente e posteriormente a esses pontos, o ônibus se encontrava na Rua Primeiro de Março, no centro da cidade, uma rua com intensa movimentação de veículos e cercada por prédios altos.

Tabela 3.4: Erro de GPS do ônibus D867990.

Data	Hora	Ordem	Linha	Lat.	Long.	Vel.	Localização
07/10/2014	16:10:42	D867990	2332	-22,91	-43,17	6,67	R. Primeiro de Março
07/10/2014	16:11:50	D867990	2332	-23,02	-43,24	3,7	Oceano
07/10/2014	16:12:25	D867990	2332	-23,02	-43,24	0,19	Oceano
07/10/2014	16:13:35	D867990	2332	-23,02	-43,24	6,67	Oceano
07/10/2014	16:14:45	D867990	2332	-23,02	-43,24	19,82	Oceano
07/10/2014	16:15:20	D867990	2332	-22,91	-43,17	8,15	R. Primeiro de Março

A Tabela 3.5 mostra a localização do ônibus de ordem A413540, que apresenta 1 localização no oceano. Anteriormente e posteriormente a esses pontos, o ônibus se encontrava na Rua San Martin, em Ipanema, uma rua residencial com árvores altas e prédios de média estatura.

Tabela 3.5: Erro de GPS do ônibus A413540.

Data	Hora	Ordem	Linha	Lat.	Long.	Vel.	Localização
07/10/2014	08:58:32	A413540	132	-22,98	-43,22	32,22	Av. San Martin x Av. Borges de Med.
07/10/2014	08:59:32	A413540	132	-23,01	-43,22	0,19	Oceano
07/10/2014	09:00:42	A413540	132	-22,99	-43,22	1,85	Av. San Martin x R. Cupertino Durão

A Tabela 3.6 mostra a localização do ônibus de ordem A410110, que apresenta 4 localizações seguidas no oceano. Anteriormente a esse ponto, o ônibus se encontrava na Avenida Rio Branco, no centro da cidade, uma avenida com intensa movimentação de veículos e cercada por prédios altos.

Tabela 3.6: Erro de GPS do ônibus A410110.

Data	Hora	Ordem	Linha	Lat.	Long.	Vel.	Localização
07/10/2014	12:54:25	A410110	170	-22,91	-43,18	2,59	Av. Rio Branco
07/10/2014	12:55:35	A410110	170	-22,9	-43,18	30,19	Av. Rio Branco
07/10/2014	12:56:45	A410110	170	-22,9	-43,18	30,19	Oceano
07/10/2014	12:57:20	A410110	170	-22,9	-43,18	30,19	Oceano
07/10/2014	12:59:13	A410110	170	-22,98	-43,15	18,33	Oceano
07/10/2014	13:00:23	A410110	170	-22,98	-43,15	25,19	Oceano
07/10/2014	13:04:31	A410110	170	-22,9	-43,19	0	Central do Brasil

A Tabela 3.7 mostra a localização do ônibus de ordem B710430, que apresenta 11 localizações seguidas no oceano. O ônibus, entre 00:07h e 05:08h, reportou a localização na mesma posição, no oceano. Em seguida, às 05:14h, o ônibus reportou a localização na garagem da empresa. O local possivelmente é coberto por um telhado, interferindo na localização dos sinais de satélite.

Tabela 3.7: Erro de GPS do ônibus B710430.

Data	Hora	Ordem	Linha	Lat.	Long.	Vel.	Localização
07/10/2014	00:07:58	B710430	455	-22,98	-43,04	0	Oceano
07/10/2014	00:37:58	B710430	455	-22,98	-43,04	0	Oceano
07/10/2014	01:07:58	B710430	455	-22,98	-43,04	0	Oceano
07/10/2014	01:37:59	B710430	455	-22,98	-43,04	0	Oceano
07/10/2014	02:07:59	B710430	455	-22,98	-43,04	0	Oceano
07/10/2014	02:37:59	B710430	455	-22,98	-43,04	0	Oceano
07/10/2014	03:07:59	B710430	238	-22,98	-43,04	0	Oceano
07/10/2014	03:37:59	B710430	238	-22,98	-43,04	0	Oceano
07/10/2014	04:08:00	B710430	238	-22,98	-43,04	0	Oceano
07/10/2014	04:38:00	B710430	238	-22,98	-43,04	0	Oceano
07/10/2014	05:08:00	B710430	238	-22,98	-43,04	0	Oceano
07/10/2014	05:14:29	B710430	238	-22.9024	-43.3148	0	Garagem, R. Torres de Oliveira
07/10/2014	05:15:45	B710430	238	-22.9024	-43.3148	0	Garagem, R. Torres de Oliveira
07/10/2014	05:16:45	B710430	238	-22.9024	-43.3148	0	Garagem, R. Torres de Oliveira

A Tabela 3.8 mostra a localização do ônibus de ordem C414150, que apresenta 27 localizações seguidas no oceano. Anteriormente a esses pontos, o ônibus se encontrava na Avenida Presidente Vargas, no centro da cidade, uma avenida com intensa movimentação de veículos e cercada por prédios altos. O ônibus reportou por 30 minutos a localização de forma errada, além disso, apresentou nesses pontos a velocidade de 100 km/h. O itinerário desta linha de ônibus passa pela zona sul da cidade e por um túnel (Túnel Zuzu Angel), locais sujeitos a interferência do sinal de GPS. A localização é novamente reportada de forma correta somente na Avenida Lúcio Costa, na Barra da Tijuca.

Tabela 3.8: Erro de GPS do ônibus C414150.

Data	Hora	Ordem	Linha	Lat.	Long.	Vel.	Localização
07/10/2014	05:57:38	C414150	309	-22,91	-43,19	18,71	Av. Presidente Vargas
07/10/2014	05:58:48	C414150	309	-22,91	-43,19	18,71	Av. Presidente Vargas
07/10/2014	05:59:23	C414150	309	-22,91	-43,19	18,71	Av. Presidente Vargas
07/10/2014	06:00:33	C414150	309	-22,91	-43,19	18,71	Av. Presidente Vargas
07/10/2014	06:01:43	C414150	309	-22,91	-43,19	18,71	Av. Presidente Vargas
07/10/2014	06:02:51	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:03:26	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:04:36	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:05:46	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:06:21	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:06:56	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:07:31	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:08:41	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:10:26	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:11:36	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:12:46	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:13:21	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:14:31	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:15:41	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:17:26	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:18:36	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:19:46	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:20:21	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:21:31	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:22:41	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:23:16	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:24:26	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:25:36	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:26:46	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:27:21	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:28:31	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:29:41	C414150	309	-22,86	-43,12	100,01	Oceano
07/10/2014	06:30:38	C414150	309	-23,01	-43,37	32,22	Av. Lúcio Costa x Ayrton Senna
07/10/2014	06:31:48	C414150	309	-23,01	-43,36	50,37	Av. Lúcio Costa x Ayrton Senna
07/10/2014	06:32:23	C414150	309	-23,01	-43,36	68,15	Av. Lúcio Costa x Ayrton Senna

A Tabela 3.9 mostra a localização do ônibus de ordem A411830, que apresenta 3 localizações seguidas no oceano. Anteriormente a esses pontos, o ônibus se encontrava na entrada do Túnel Zuzu Angel e posteriormente a localização foi reportada a 500 metros da saída do mesmo túnel.

Tabela 3.9: Erro de GPS do ônibus A411830.

Data	Hora	Ordem	Linha	Lat.	Long.	Vel.	Localização
07/10/2014	12:25:42	A411830	178	-22,99	-43,25	30	Autoestrada Lago-Barra (Entrada Túnel)
07/10/2014	12:27:42	A411830	178	-22,99	-43,25	26,67	Autoestrada Lago-Barra (Entrada Túnel)
07/10/2014	12:28:43	A411830	178	-22,99	-43,25	31,85	Autoestrada Lago-Barra (Entrada Túnel)
07/10/2014	12:29:43	A411830	178	-22,99	-43,25	31,85	Autoestrada Lago-Barra (Entrada Túnel)
07/10/2014	12:30:48	A411830	178	-23,03	-43,3	89,27	Oceano
07/10/2014	12:31:54	A411830	178	-23,03	-43,3	39,63	Oceano
07/10/2014	12:32:33	A411830	178	-23,03	-43,3	17,78	Oceano
07/10/2014	12:33:28	A411830	178	-22,98	-43,22	26,85	Rua Mário Ribeiro (Saída Túnel)
07/10/2014	12:34:53	A411830	178	-22,98	-43,22	0	Rua Mário Ribeiro (Saída Túnel)

A Tabela 3.10 mostra a localização do ônibus de ordem A480840, que apresenta 3

localizações seguidas na Floresta da Tijuca, local de mata fechada sem nenhuma via de circulação de veículos. Anteriormente a esses pontos, o ônibus se encontrava na Rua Frei Caneca, no centro da cidade, uma rua com árvores, intensa movimentação de veículos e cercada por prédios altos.

Tabela 3.10: Erro de GPS do ônibus A480840.

Data	Hora	Ordem	Linha	Lat.	Long.	Vel.	Localização
07/10/2014	08:52:37	A480840	201	-22,92	-43,2	0	R. Frei Caneca x Riachuelo
07/10/2014	08:53:19	A480840	201	-22,91	-43,2	22,41	R. Frei Caneca x Riachuelo
07/10/2014	08:54:26	A480840	201	-22,91	-43,2	34,82	R. Frei Caneca x Riachuelo
07/10/2014	08:55:02	A480840	201	-22,91	-43,19	59,45	R. Frei Caneca x Riachuelo
07/10/2014	08:59:39	A480840	201	-22,94	-43,3	1,11	Floresta Tijuca
07/10/2014	09:00:25	A480840	201	-22,94	-43,3	1,85	Floresta Tijuca
07/10/2014	09:01:00	A480840	201	-22,94	-43,3	0	Floresta Tijuca
07/10/2014	09:02:10	A480840	201	-22,94	-43,3	3,7	Floresta Tijuca
07/10/2014	09:03:20	A480840	201	-22,91	-43,18	16,3	Av. República do Chile
07/10/2014	09:04:30	A480840	201	-22,91	-43,18	6,85	Av. República do Chile
07/10/2014	09:05:40	A480840	201	-22,91	-43,18	15	Av. República do Chile
07/10/2014	09:06:15	A480840	201	-22,91	-43,18	0	Av. República do Chile
07/10/2014	09:07:25	A480840	201	-22,91	-43,18	2,04	Av. República do Chile
07/10/2014	09:08:36	A480840	201	-22,91	-43,18	2,96	Av. República do Chile
07/10/2014	09:09:30	A480840	201	-22,91	-43,18	11,48	Av. República do Chile
07/10/2014	09:10:38	A480840	201	-22,91	-43,18	3,33	Av. República do Chile

Os erros identificados a partir do levantamento são ocasionados em locais propensos a erros de recepção e/ou a interferências do sinal do GPS no equipamento receptor do veículo. Nesses locais de interferência, quando o veículo está trafegando entre prédios elevados ou abaixo de árvores, podem acontecer reflexões, atenuações ou até mesmo bloqueio em uma determinada direção da recepção de sinais de GPS, dificultando a localização da quantidade necessária de satélites distintos para determinar a posição do veículo com precisão.

Como possíveis correções existe a possibilidade de remover os locais indicados de forma errônea, como áreas do oceano ou florestas, ou a identificação da possível trajetória, inserindo a posição do veículo nesses intervalos de tempo. Além disso, os locais dos erros não são conhecidos, ou seja, não é informado a precisão do dado coletado, não sendo imediata a localização dos erros.

Dessa forma, como a quantidade de erros de localização observados não é significativa já que poucas ocasiões apresentam erros e esses erros não são reportados como uma única posição, e sim espalhados em diversos locais, principalmente no oceano ou ao redor das vias em que trafegam, não interferindo nas análises realizadas nesta tese já que nenhum cluster é formado sobre áreas desses tipos (pontos hidrográficos e florestas).

3.2.4 Análise dos padrões de deslocamento dos ônibus

Dando continuidade à análise, o primeiro dado a ser observado são as quantidades de entradas de dados, aqui chamadas também de atualizações, presentes no banco

de dados em determinado período de tempo. Na Figura 3.2 é possível observar o número de atualizações enviadas a cada hora durante o dia 07 de outubro de 2014. O maior número de atualizações ocorre das 05:00h às 20:00h, próximas ao horário comercial, que é das 08:00h às 17:00h, o que é um comportamento esperado, pois é o período que mais se utiliza o transporte público e portanto o período em que há mais ônibus circulando.

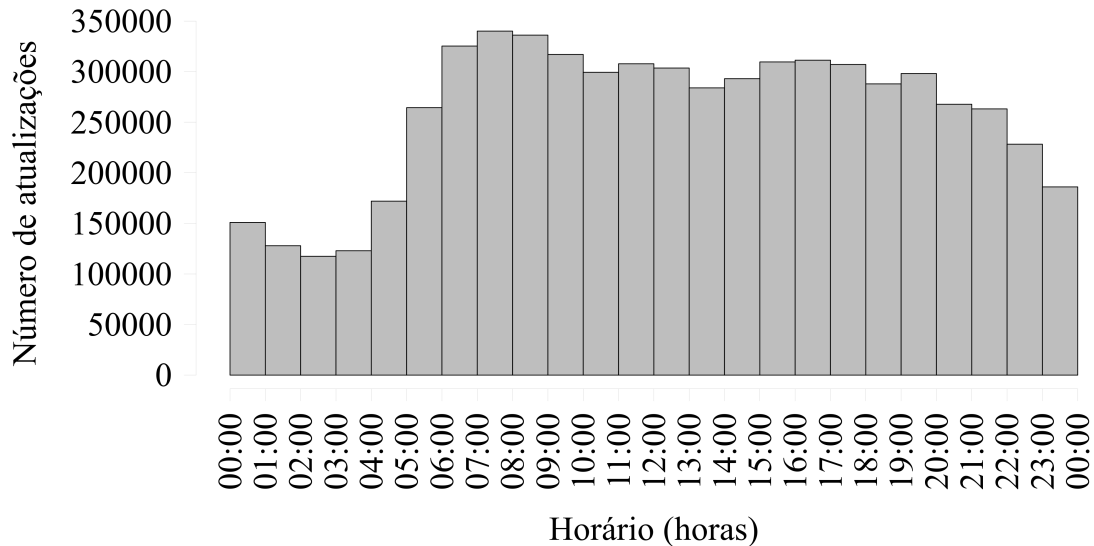


Figura 3.2: Atualizações dos ônibus durante 24 horas do dia 07/10/2014.

A mesma observação de atualizações pode ser feita para o período de uma semana. Na Figura 3.3 é possível observar o mesmo comportamento de atualizações durante o horário comercial, entretanto, no final de semana, dias 11 e 12 de outubro de 2014, o comportamento das atualizações é diferente. O número máximo de atualizações é reduzido consideravelmente, entretanto o comportamento de maior número de atualizações no horário da manhã e da tarde é mantido. É possível observar também uma falta de dados entre os dias 9 e 10 de outubro, que foi ocasionada pela indisponibilidade do sistema de localização dos ônibus pela prefeitura do Rio de Janeiro.

A Figura 3.4 mostra a quantidade de ônibus que tem uma determinada frequência diária de atualizações. O gráfico apresenta no eixo X a quantidade de atualizações presentes no banco de dados para cada veículo e no eixo Y quantos veículos têm determinado número de atualizações de localização durante as 24 horas. O número máximo de atualizações em um dia é 1.440, porém poucos ônibus atualizam a sua posição nessa frequência. Uma grande parte dos ônibus atualiza a sua posição por volta de 900 vezes ao dia, o que representa 15 horas de atualizações de dados. Esse dado mostra que poucos ônibus reportam a sua posição durante as 24 horas do dia e que muitos ônibus reportam a posição poucas vezes durante o dia, o que pode acontecer por razões diversas como ônibus que não transmitem a posição quando

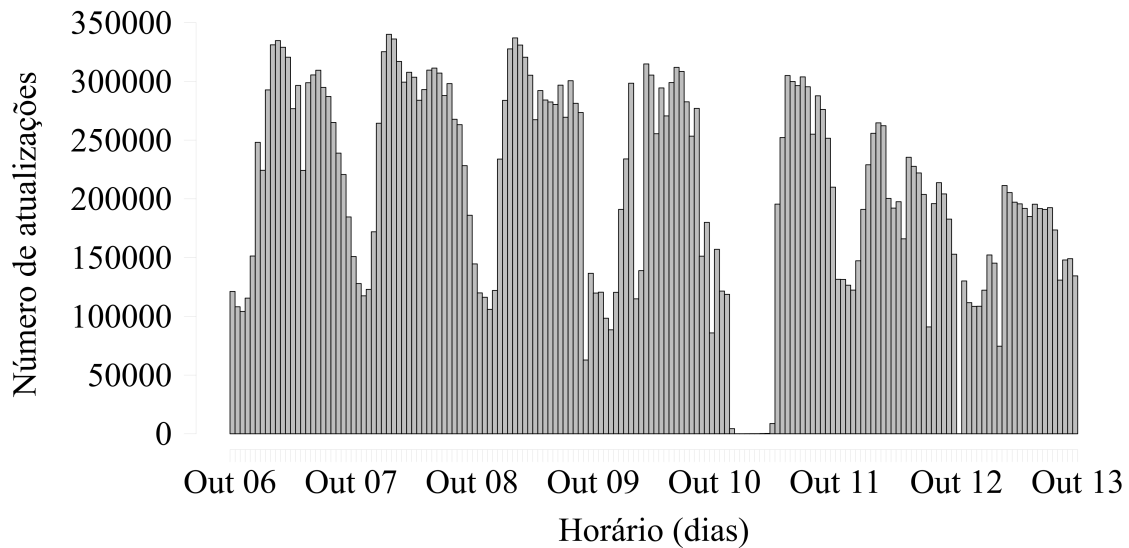


Figura 3.3: Atualizações dos ônibus durante a semana de 06 a 12/10/2014.

desligados, falhas de transmissão, alguns dados de localização não sendo armazenados no banco de dados da prefeitura, etc.

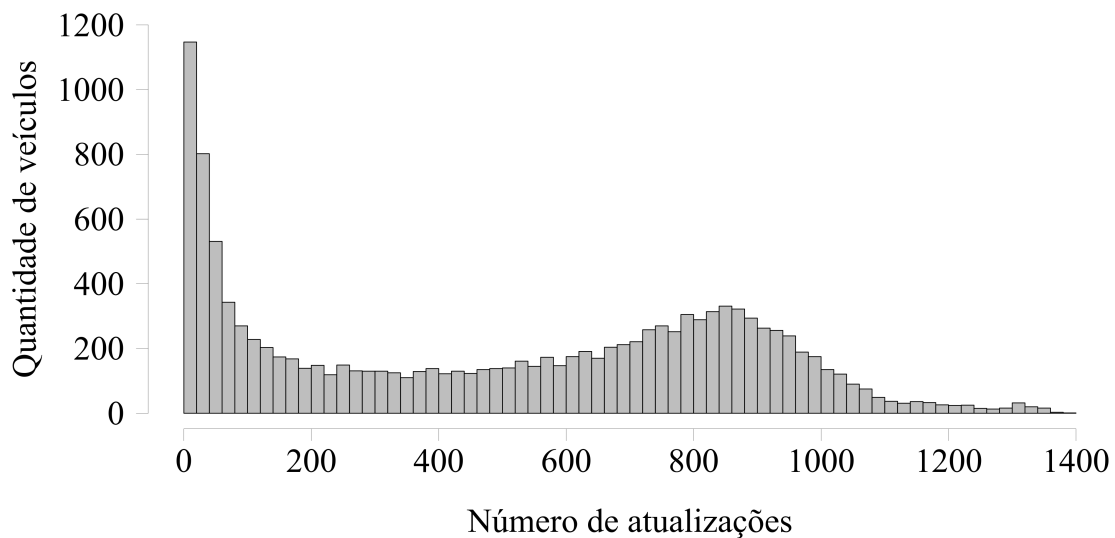


Figura 3.4: Frequência de atualizações diárias por ônibus.

Outra característica é o intervalo entre dados de localização dos ônibus. A Figura 3.5 apresenta no eixo da esquerda um histograma com as informações de intervalo de atualização medidas no banco de dados. A maior concentração é de intervalos de atualização próximos a um minuto, porém nunca é menor do que 1 minuto. Já o gráfico de linha representa a probabilidade cumulativa (CDF) dos intervalos de atualização. Observa-se que 90% dos intervalos de atualização são menores que 7 minutos, ou seja, é necessário que se considere esse intervalo entre atualizações para obter a maioria dos dados de localizações dos ônibus. Essa figura mostra que os dados de localização enviados não tem um intervalo constante, não podendo ser

considerados homogêneos ou sincronizados. Aqui vale mencionar que estes intervalos de atualização, em tese, deveriam ser constantes e iguais a 1 minuto, intervalo previsto entre as atualizações enviadas pelos ônibus e que está em desacordo com o indicado pela mantenedora dos dados (Fetranspor).

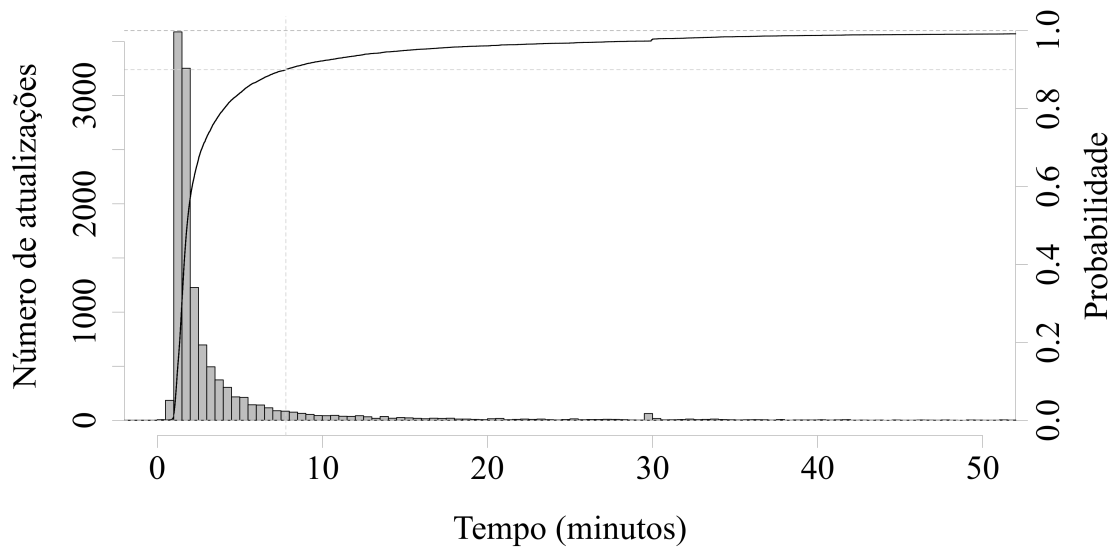


Figura 3.5: Tempo médio de intervalo de atualização dos ônibus.

Dados das velocidades reportadas pelo GPS de cada ônibus também foram analisados. Na Figura 3.6 é exibido o histograma dos valores de velocidade informados pelos ônibus durante um dia. A Figura 3.7 mostra um histograma com os mesmos dados de velocidade porém com uma divisão maior de faixas de velocidade. A representação do valor da velocidade de 0 km/h a 5 km/h foi limitada, pois é consideravelmente maior do que os outros valores.

A velocidade máxima permitida para ônibus é de 90 km/h em rodovias e de 80 km/h em perímetro urbano [57]. É possível observar que pelo menos 14.551 (0,2%) de dados do GPS informavam velocidades acima da permitida, o que pode ser ocasionado devido a erros de GPS (Seção 3.2.3) ou por excesso de velocidade. Além disso, a maior parte dos dados de velocidade está concentrada em velocidades baixas, de até 20 km/h, o que mostra que os ônibus tem a tendência de trafegar em baixa velocidade. O fato de ser necessário pegar passageiros durante o trajeto, necessitando de paradas para embarque e desembarque e a presença de sinais de trânsito no perímetro urbano, acarreta em uma elevada quantidade de ônibus reportando a velocidade até 5 km/h. Além disso, esta tendência de baixas velocidade também pode estar relacionada ao trânsito da cidade e às características dos ônibus, que não são construídos para transitar em alta velocidade.

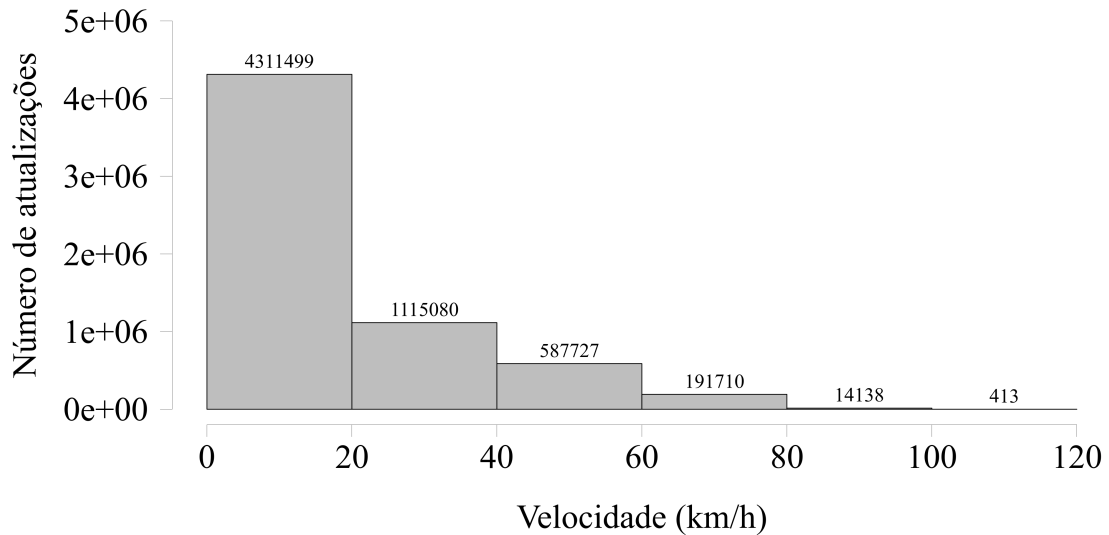


Figura 3.6: Velocidade reportada pelo GPS dos ônibus.

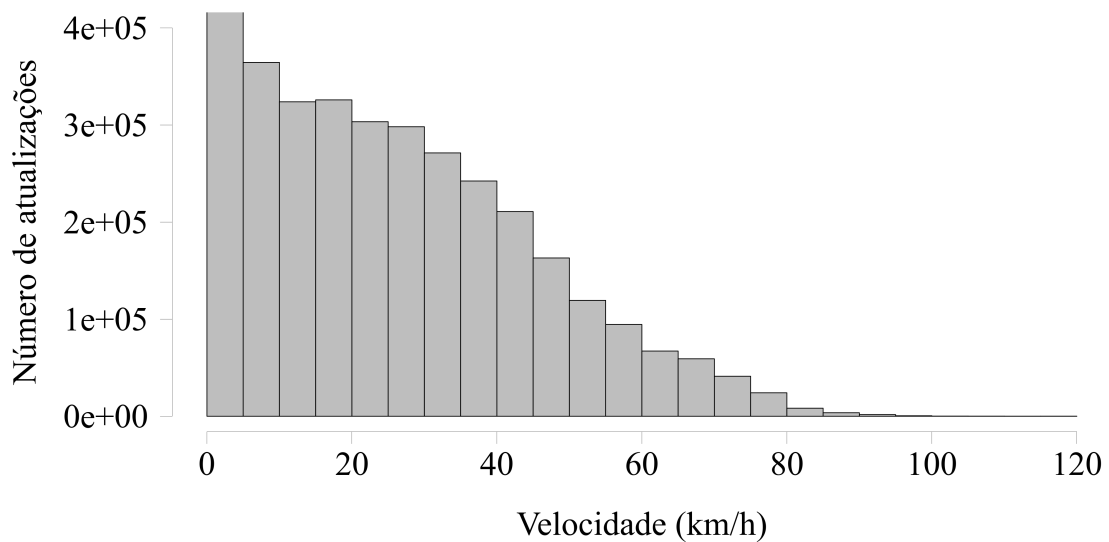


Figura 3.7: Detalhamento da velocidade reportada pelo GPS dos ônibus.

Na Figura 3.8 é exibido um mapa com as posições dos ônibus entre 07:00h e 07:05h de um dia. As cores que representam a velocidade são graduadas começando no vermelho, com velocidades entre 0-18 km/h; passando pelo laranja, com velocidades entre 18-36 km/h; o amarelo representado as velocidades entre 36-53 km/h; o verde-claro representando as velocidades entre 53-71 km/h; e o verde-escuro para velocidades entre 71-89 km/h. O mapa é dominado pelas velocidades baixas, entretanto, em algumas vias expressas da cidade, como na, Avenida Brasil, Linha Amarela, Linha Vermelha, Aterro do Flamengo e Avenida das Américas (indicadas pelas setas) é possível observar maiores velocidades. Nessas vias existem poucos ou nenhum ponto de ônibus para pegar passageiros. A maior concentração de pontos vermelhos ocorre no centro da cidade, local que contém uma grande quantidade de

pontos de ônibus e de outros veículos no horário observado.

Os valores podem ser utilizados como métrica de congestionamento da cidade. Entretanto, é necessário ressaltar que os ônibus necessariamente precisam parar nos pontos de ônibus para permitir o fluxo de passageiros. Além disso, a cidade do Rio de Janeiro implementa o sistema de BRS (*Bus Rapid Service* ou Serviço Rápido por Ônibus), com faixas exclusivas para ônibus, o que diminui o tempo de deslocamento dos ônibus em até 20% [58].

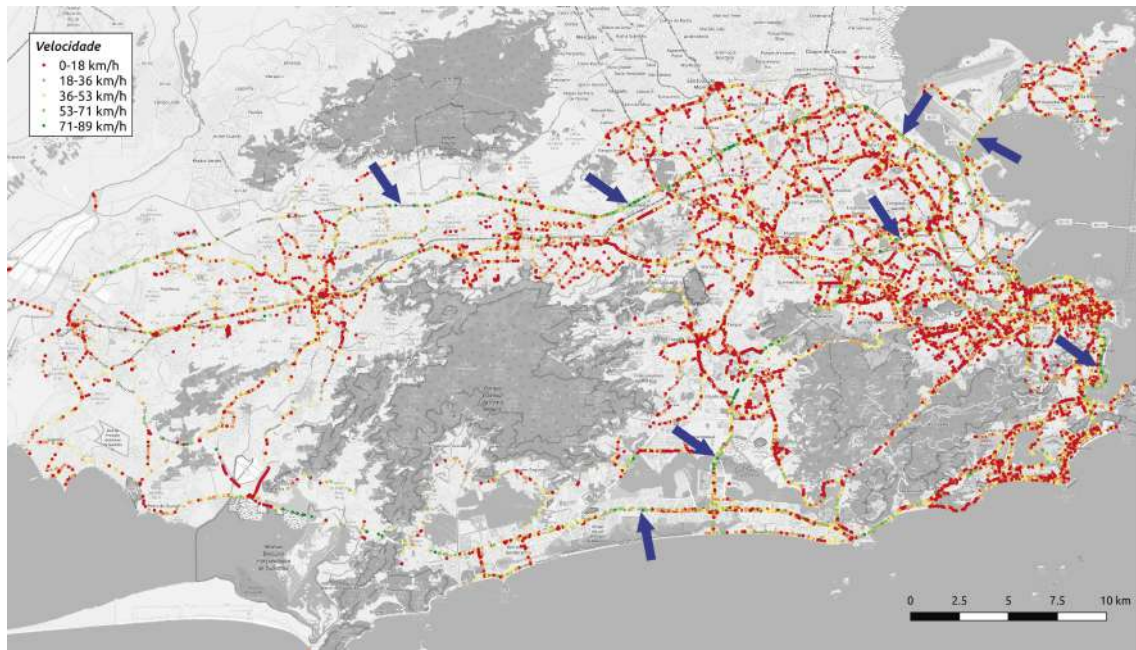


Figura 3.8: Mapa com a velocidade dos ônibus entre 07:00h e 07:05h.

Outro dado analisado é a quantidade de linhas distintas que um ônibus pode ter, ou seja, a quantidade de mudanças de linhas que ocorrem em um dia para cada veículo. Algumas entradas do banco de dados não trazem informações da linha, constando zero como o seu valor. Essas informações foram desconsideradas na Figura 3.9, onde é exibido o número de linhas que um ônibus assume durante um dia. Grande parte dos ônibus faz a mesma linha durante todo o dia. Porém, alguns (poucos) ônibus assumem mais de uma linha por dia. Somente 94 de 12.456 ônibus tiveram 4 ou mais linhas diferentes em um dia.

A Figura 3.10 mostra o número de ônibus que informam ao banco de dados que estão designados àquela linha. Grande parte das linhas tem menos de 50 ônibus distintos no período de um dia. Esse número não significa que esses ônibus estão na mesma linha ao mesmo tempo, existe a possibilidade da troca de linhas. Linhas com grandes quantidades de ônibus normalmente são as que transportam a maior quantidade de passageiros. A quantidade de ônibus por linha é definida em contrato.

A partir da análise de dados pode-se evidenciar alguns padrões presentes no sistema de transporte da cidade do Rio de Janeiro. É possível determinar as lo-

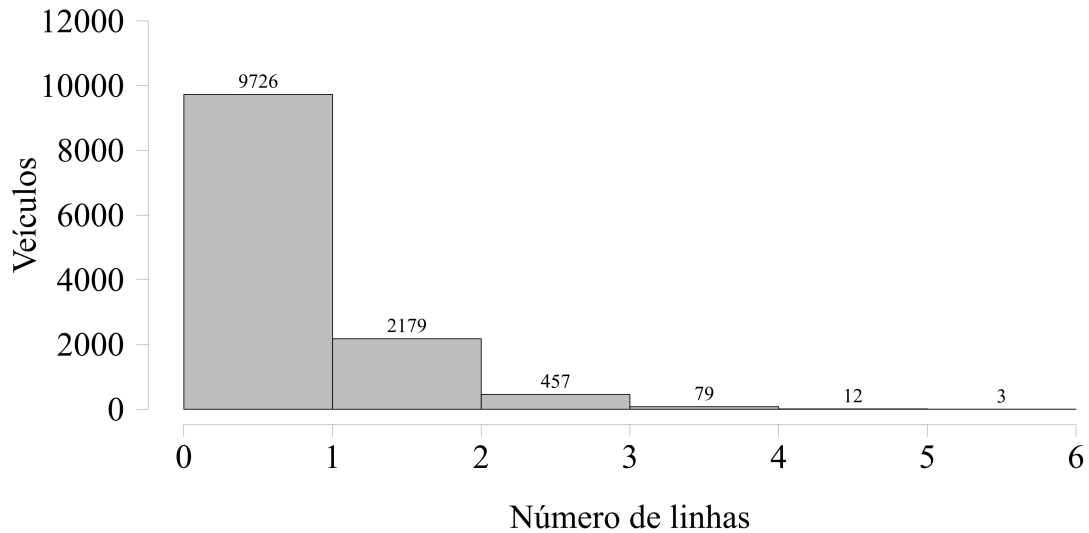


Figura 3.9: Número de linhas de um ônibus durante um dia.

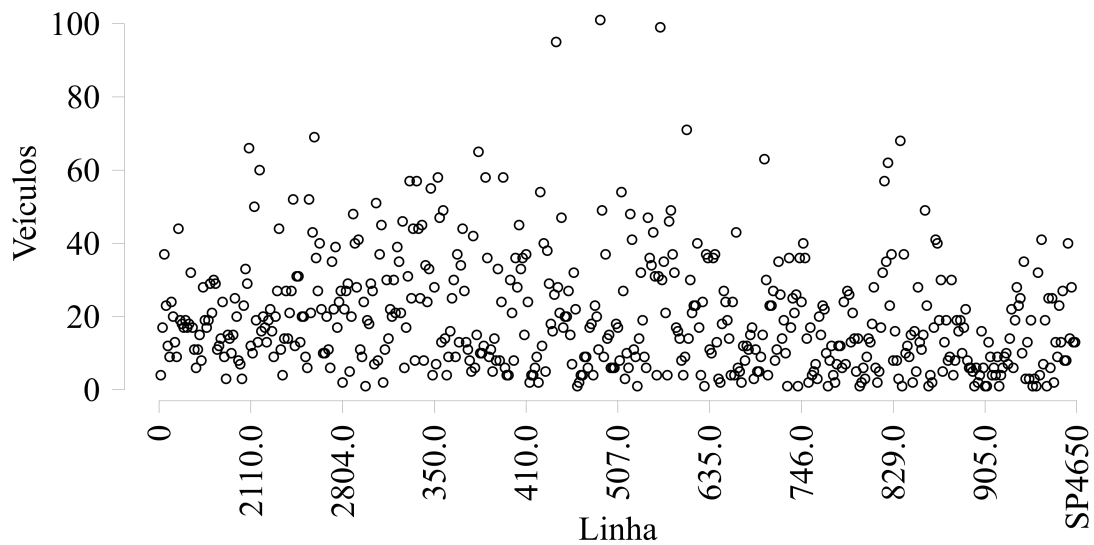


Figura 3.10: Número de ônibus distintos que estão designados à linha.

calizações em que os ônibus permanecem por um período maior de tempo, como pontos finais, garagens, áreas de tráfego intenso, assim como locais onde há um número elevado de ônibus durante um período de tempo, como pontos nas centrais e pontos de transbordo. Esses dados são utilizados para planejar a distribuição da infraestrutura necessária para operar a rede proposta nesta tese e para determinar a capacidade de dados que podem ser transportados, como será apresentado no decorrer desta tese, além de serem importantes para a prefeitura da cidade no planejamento e organização da rede de transporte.

3.3 Dados de posicionamento dos táxis

Conforme mencionado anteriormente, os dados de posicionamento dos táxis aqui apresentados foram cedidos após o contato com a empresa 99 (antiga 99Taxi). Os dados são coletados através de dispositivos móveis (*smartphones*) de posse de cada um dos motoristas de táxi que utilizam o aplicativo 99 na cidade do Rio de Janeiro. O taxista é registrado na 99, instala o aplicativo em seu *smartphone* e o executa para conseguir atender às requisições de passageiros. As requisições de viagens, de forma semelhante, são efetuadas por um outro aplicativo instalado no *smartphone* dos passageiros.

A empresa 99 é então responsável por conectar pessoas que desejam efetuar uma viagem de táxi a taxistas que desejam transportar essa pessoa. Para isso, é essencial a localização de cada cliente e a localização de cada um dos taxistas próximos. A partir de uma solicitação à 99, foram disponibilizadas, para esse estudo, as localizações de cada um dos taxistas registrados no dia 07 de outubro de 2014. A empresa informou que os dados não apresentam tempo de atualização definido, podendo ser tão curtos quanto vários dados de localização por minuto. O Código 3.2 mostra um exemplo dos dados de localização de um dos táxis monitorados. Além disso, no ano de 2014, a empresa ainda não possuía transporte realizado por carros particulares, ou seja, os dados são somente de táxis legalizados e cadastrados no aplicativo.

Código 3.2: Parte dos dados recebidos da 99

```
locationDate , latitude , longitude
2014-10-07 13:57:01 , -22.9085657 , -43.19786555
2014-10-07 14:07:15 , -22.90556948 , -43.19075568
2014-10-07 14:07:48 , -22.9055696 , -43.19075495
2014-10-07 14:07:02 , -22.90557376 , -43.19074238
2014-10-07 14:10:16 , -22.90349994 , -43.19076276
2014-10-07 14:07:15 , -22.90556948 , -43.19075568
2014-10-07 14:10:59 , -22.90298016 , -43.19103537
2014-10-07 14:13:06 , -22.90102432 , -43.19271637
2014-10-07 14:13:12 , -22.90040339 , -43.19329844
2014-10-07 14:12:56 , -22.90176278 , -43.19201278
2014-10-07 14:14:50 , -22.89711564 , -43.1912347
2014-10-07 14:12:14 , -22.90207997 , -43.19174581
2014-10-07 14:16:44 , -22.8986267 , -43.18762674
2014-10-07 14:17:25 , -22.9016949 , -43.18506432
```

3.3.1 Análise dos padrões de deslocamento dos táxis

Assim como o banco de dados de ônibus, o banco de dados de táxis possui características únicas sobre a coleta de dados de posicionamento dos táxis monitorados. Por questões de segredo empresarial, o código fonte e o funcionamento preciso do programa de monitoramento das posições dos táxis não está disponível para consulta. Dessa forma, as características de atualização de posição são distintas das dos ônibus. Entretanto, a empresa informou que a localização dos táxis é informada à central através de qualquer conexão de Internet que o celular do taxista contenha, seja ela IEEE 802.11 ou 3G/4G. Assim sendo, fatores que afetam a transmissão da informação de localização são semelhantes, como o congestionamento ou falta de cobertura da rede de celular, que afeta o envio de informações de localização dos ônibus.

A Figura 3.11 mostra todos os dados coletados no dia 07 de outubro de 2014, onde cada ponto representa um dado de posicionamento de um táxi recebido nesse dia. É possível observar a presença de táxis em todas as regiões da cidade, porém de menor forma porção esquerda da figura. Locais típicos de táxis, como aeroportos, tem uma elevada quantidade de pontos. Também são observados erros de localização, como pontos no oceano e em regiões de floresta, porém a quantidade é menor do que os erros contidos no banco de dados dos ônibus.

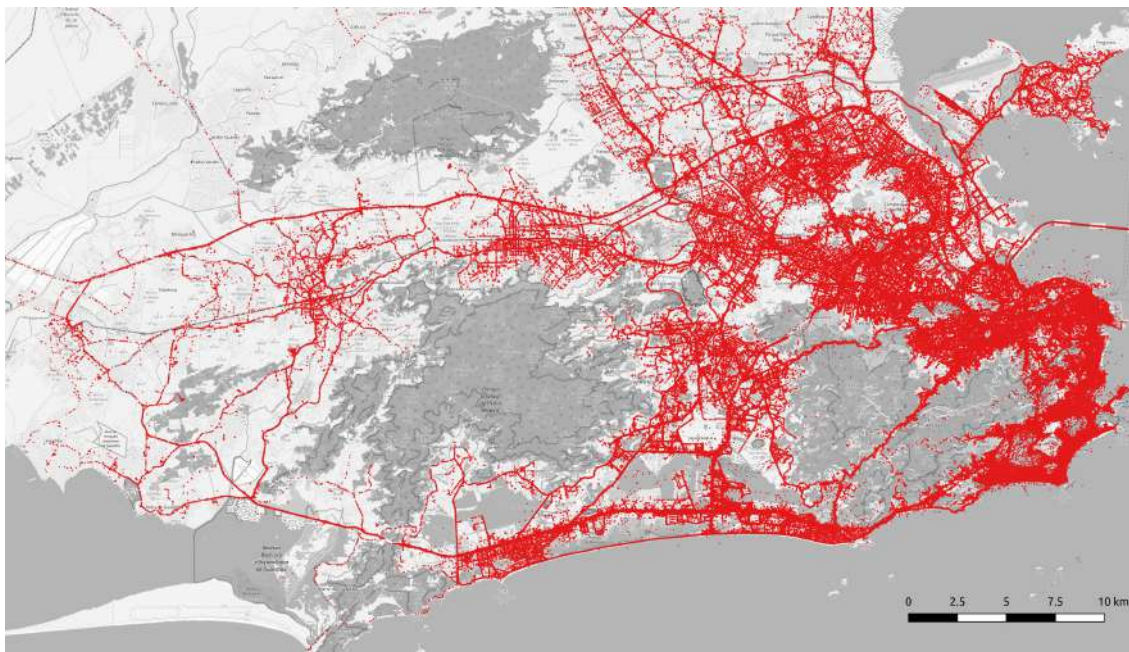


Figura 3.11: Mapa com a localização dos táxis no dia 07/10/2014.

O primeiro dado a ser observado são as quantidades de dados de posicionamento, ou atualizações, que estão presentes no banco de dados da empresa 99. Na Figura 3.12, é possível observar um pico na quantidade de atualizações durante a madrugada, das 02:00h às 04:00h, e uma elevada quantidade de atualizações durante

o dia, a partir de 08:00h até 23:59h.

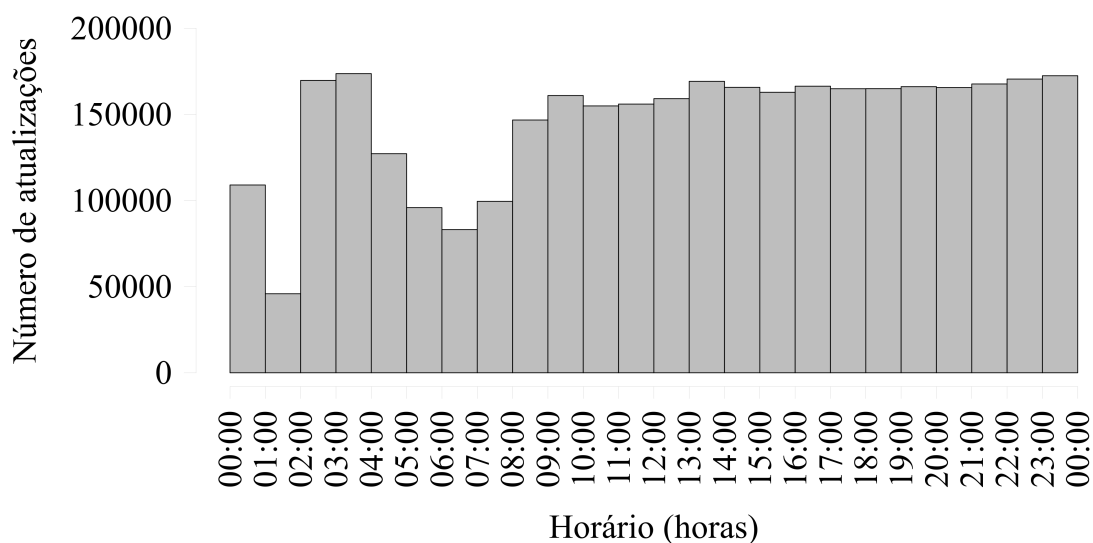


Figura 3.12: Atualizações dos táxis durante as 24 horas do dia 07/10/2014.

A Figura 3.13 mostra a quantidade de táxis com uma determinada frequência diária de atualizações. Atualizações a cada minuto geram 1.440 entradas de dados por dia. O gráfico apresenta no eixo X a quantidade de atualizações presentes no banco de dados para cada veículo e no eixo Y quantos veículos têm determinado número de atualizações de localização durante as 24 horas. É possível observar que existem alguns táxis que enviam a sua posição mais de uma vez por minuto.

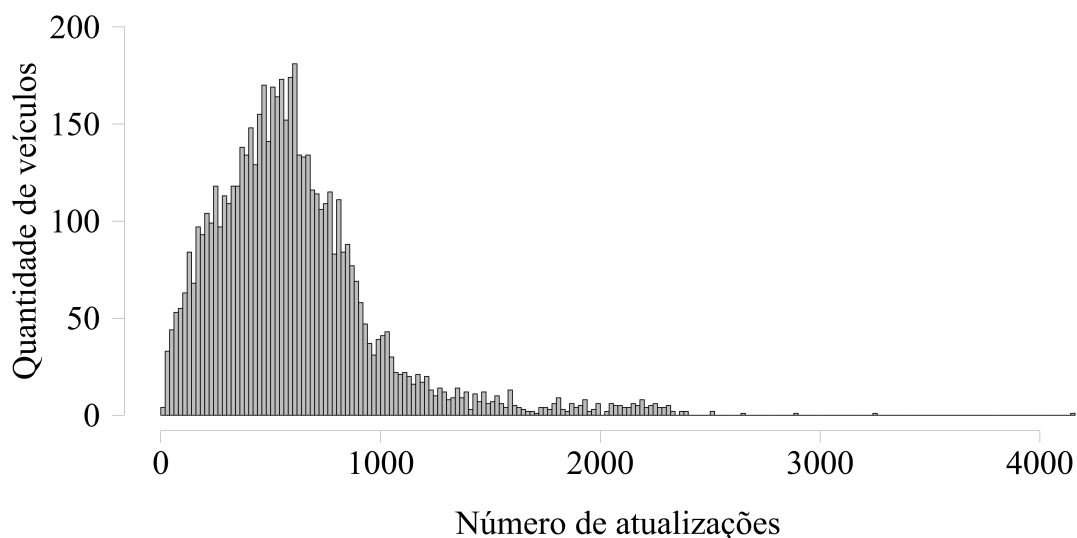


Figura 3.13: Frequência de atualizações diárias por táxis.

Outra característica importante é o intervalo entre dados de localização dos táxis, já que esse é um dado não informado pela 99, sendo essencial para determinar o tipo de análise que pode ser realizado com os dados. A Figura 3.14 apresenta no eixo da direita um histograma com as informações de intervalo de atualização médio no

banco de dados. A maior concentração é de intervalos de atualização próximos a 1 minuto. Já o gráfico de linha representa a probabilidade cumulativa dos intervalos de atualização. Observa-se que 90% dos intervalos de atualização são menores que 3 minutos.

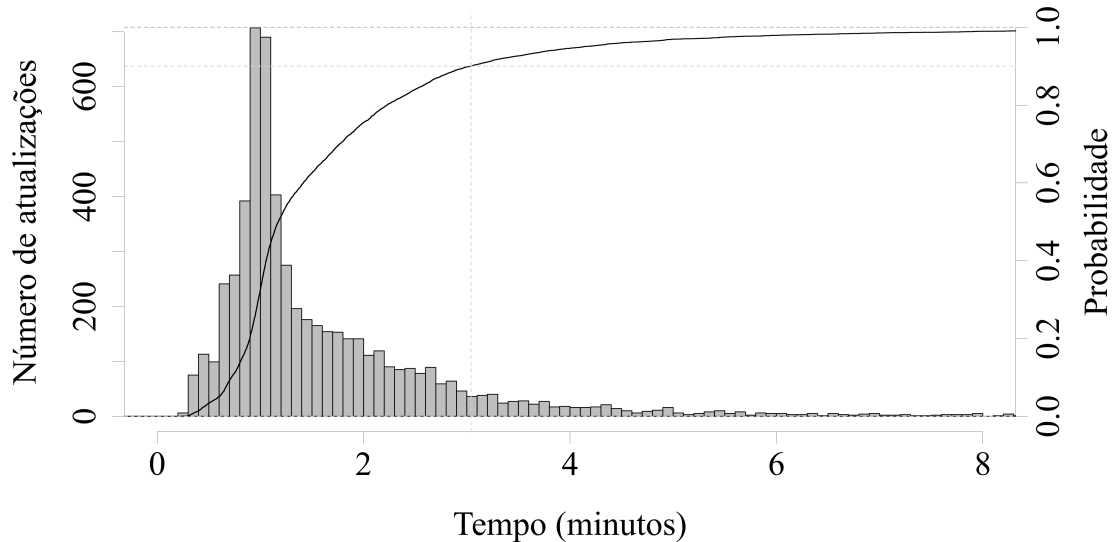


Figura 3.14: Tempo médio de intervalo de atualização dos táxis.

3.3.2 Análise dos padrões de deslocamento dos veículos públicos da cidade do Rio de Janeiro

A partir da junção dos dois bancos de dados citados, de ônibus e de táxis, é possível obter uma análise conjunta, com uma quantidade de dados mais significativa e com capacidade de mostrar de forma mais fiel a realidade do transporte público da cidade do Rio de Janeiro. Além disso, algumas das características que uma rede veicular mista teria, diferentes das redes de ônibus ou de táxis exclusivamente, podem ser investigadas.

O arquivo de dados de localização dos táxis foi formatado para ter o mesmo formato do arquivo de ônibus, onde a velocidade não foi determinada, sendo considerada 0, a informação da linha também não é considerada e o número de ordem são as identificações dos táxis anonimizadas fornecidas pela 99. O arquivo produzido pela fusão dos dados do ônibus e dos táxis contém 9.754.014 entradas, com 18.259 identificadores de veículos distintos. O arquivo contém 675 MB e representa as posições reportadas no dia 07 de outubro de 2014.

A Figura 3.15 mostra todos os dados coletados no dia 07 de outubro de 2014, onde cada ponto representa um dado de posicionamento de cada veículo recebido nesse dia. Os pontos pretos representam os ônibus e os pontos vermelhos representam os

táxis. A circulação de táxis não se restringe à cidade do Rio de Janeiro, observando-se alguns veículos em cidades vizinhas.

Além disso, é possível observar que os ônibus estão presentes nas vias mais importantes da cidade, principalmente nas que ligam suas diversas regiões. Os táxis têm grande presença em vias secundárias. A distribuição por regiões da cidade é desigual, nas áreas de menor IDH [59] (Índice de Desenvolvimento Humano) existe de pouca a nenhuma presença de táxis que utilizam o aplicativo 99 (parte superior esquerda do mapa). Nas regiões de maior IDH existe a maciça presença tanto de táxis como de ônibus, como exemplificado na região sul da cidade (parte inferior direita do mapa).

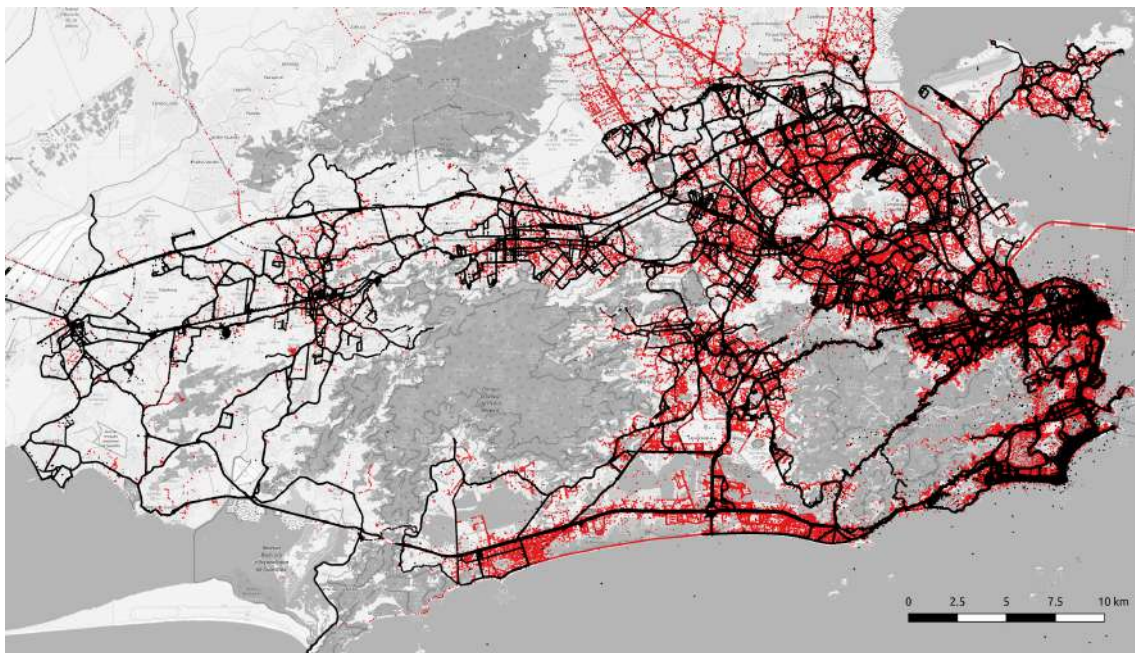


Figura 3.15: Mapa com localizações de ônibus e táxis as localizações no dia 07/10/2014.

O primeiro dado observado é a quantidade de dados de posicionamento, ou atualizações, presentes no banco de dados. A Figura 3.16 mostra, na forma de barras empilhadas, em azul as quantidade de atualizações dos ônibus e em vermelho a quantidade de atualizações dos táxis. A tendência permanece a mesma, com um período de atividade elevado durante o horário comercial. Entretanto, a quantidade de atualizações nos horários de 00:00h até 05:00h é aumentada ao introduzir os dados dos táxis. Também é possível analisar a diferença de atualizações entre os ônibus e os táxis, onde os táxis apresentam uma maior circulação do que os ônibus entre 02:00h e 04:00h.

A Figura 3.17 mostra em verde o número de atualizações de ônibus e táxis de forma conjunta, em azul, o número diário de atualizações dos ônibus e, em vermelho, o número diário de atualizações dos táxis. O gráfico apresenta no eixo X a quantidade de atualizações presentes no banco de dados para cada veículo e no eixo Y quantos

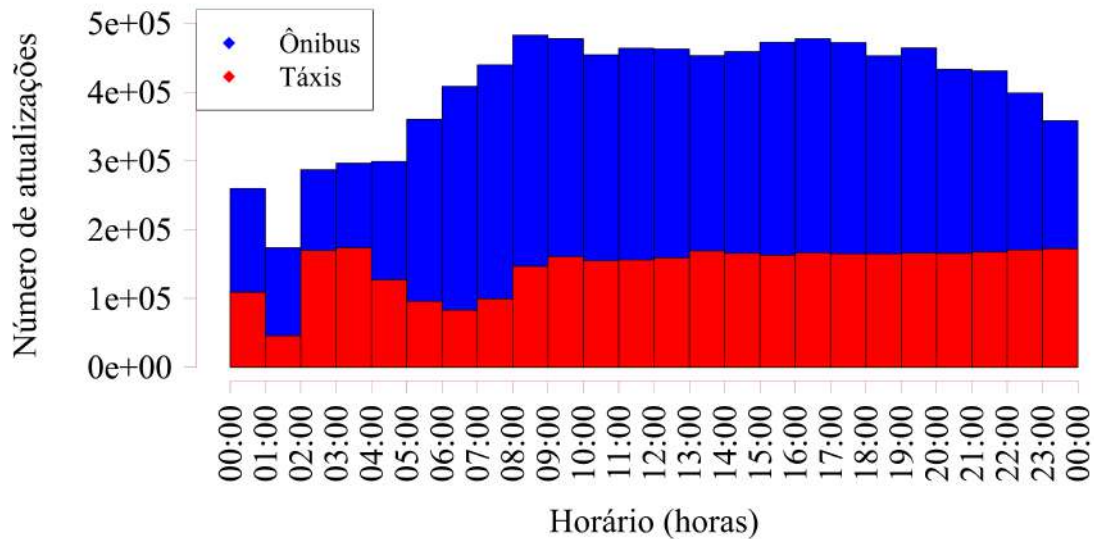


Figura 3.16: Atualizações de ônibus e táxis durante as 24 horas do dia 07/10/2014.

veículos têm determinado número de atualizações de localização durante as 24 horas. O número de veículos que envia a sua posição até 50 vezes ao dia (primeira barra) é elevado, resultante do não armazenamento de atualizações no banco de dados ou somente da baixa utilização daquele veículo no dia. Não é possível diferenciar esses dois casos com as informações obtidas. Quando se analisa ambos os veículos, grande parte das atualizações ocorrem entre 700-1000, com pico em 800 atualizações por dia.

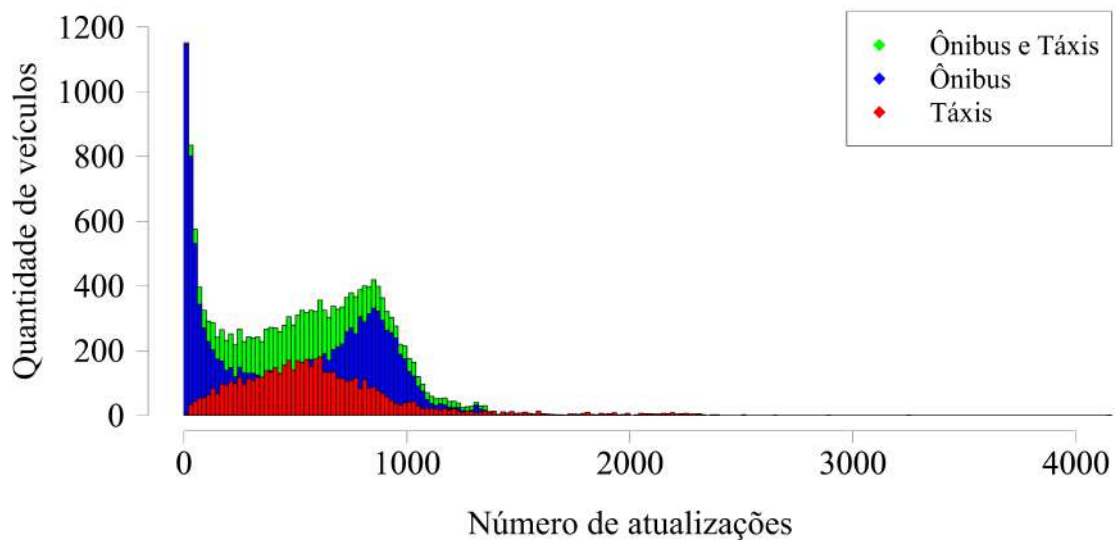


Figura 3.17: Frequências de atualizações diárias.

Outra característica importante é o intervalo entre dados de localização dos veículos. A Figura 3.18 apresenta, no eixo da direita, um histograma com as informações de intervalo de atualização médio no banco de dados. Ao juntar os dois bancos de dados, as tendências de taxas de atualizações reduzidas dos táxis fez com

que a janela de observação para encontrar 90% dos veículos fosse modificada para 6 minutos. Uma janela de observação de 5 minutos é capaz de encontrar mais de 85% dos veículos. A maior quantidade de atualizações, devido à alta representatividade dos dados dos ônibus, permanece em torno de 1 minuto.

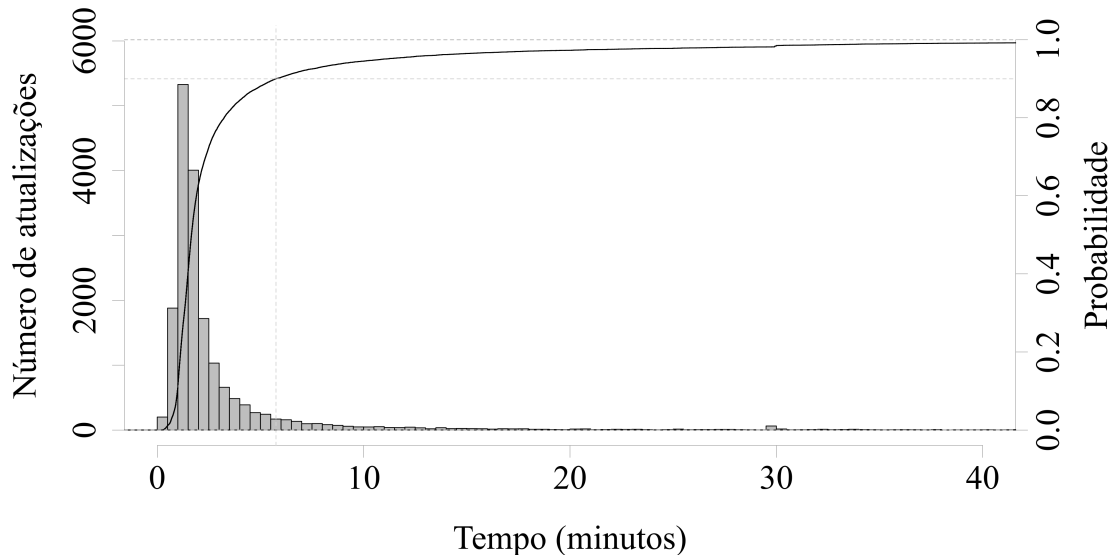


Figura 3.18: Tempo médio de intervalo de atualização de ônibus e táxis simultaneamente.

É possível também realizar a análise dos dados ao longo dos dias e ao longo da semana. O tráfego nas horas mais movimentadas do dia, horário de chegada ao trabalho e horário de saída do trabalho (08:00h e 17:00h, respectivamente) apresenta padrões distintos. Durante a manhã há a ocupação do centro da cidade com os trabalhadores vindo da periferia e durante a noite esses trabalhadores voltam para suas residências. Além dessa alteração ao longo de um dia, há diferenças entre os dias da semana. Os veículos em um final de semana não apresentam o mesmo comportamento dos veículos em um dia de semana.

3.4 Conclusão

Este capítulo apresentou a primeira análise dos bancos de dados coletados. É identificado que é possível realizar a análise ao levar em consideração as limitações encontradas. As informações coletadas são importantes para confirmar a possibilidade da realização de uma análise sobre os dados coletados, observando limitações como erros noas localizações e a taxa de atualização das posições dos ônibus a cada minuto e de forma irregular nos táxis. Além disso, é identificado o monitoramento inconstante dos veículos durante as 24 horas analisadas, levando a impossibilidade de detectar a localização dos veículos durante alguns períodos do dia.

O Capítulo 4, baseado nessas limitações, discorre sobre as possibilidades de

análise e mostra a metodologia escolhida para analisar a capacidade da rede formada por esses veículos.

Capítulo 4

Metodologia

Este capítulo objetiva apresentar as metodologias utilizadas para a análise dos dados de deslocamento de veículos do sistema de transporte público apresentados no Capítulo 3. Seu intuito é expor as ferramentas necessárias para a análise dos questionamentos desta tese, a saber:

- Como as regiões da cidade estão interligadas em relação ao deslocamento dos veículos?;
- Quanta informação é possível transmitir entre essas regiões em função do tempo?.

Dessa forma, este capítulo está dividido em quatro partes: a Seção 4.1 trata da análise quantitativa dos dados, a Seção 4.2 das ferramentas para análise de grandes massas de dados, a Seção 4.3 dos métodos de transmissão de dados e a Seção 4.4 do algoritmo de cálculo de viagens.

A análise quantitativa dos dados aborda questões relacionadas à avaliação da capacidade de transmissão utilizando os dados de localizações coletados. É realizado um estudo preliminar para avaliar a possibilidade de localizar todos os pontos da cidade propensos a transmissões de dados. Ao determinar que a quantidade de dados a ser analisados é elevada, é necessário buscar ferramentas capazes de lidar com grandes quantidades de dados. Com isso, um algoritmo de clusterização por grade é selecionado, permitindo a análise dos dados de localização dos veículos em tempo hábil. A partir do método escolhido, é necessário identificar os métodos de transmissões de dados que podem ser utilizados no cenário proposto. Finalmente, os métodos e ferramentas abordados neste capítulo permitem desenvolver um algoritmo para identificar e contabilizar os deslocamentos, ou viagens, realizados pelos veículos.

4.1 Análise quantitativa dos dados

Esta seção apresenta a primeira análise realizada a partir do banco de dados de ônibus coletado, baseada nas limitações apresentadas no Capítulo 3, como atualização de dados de localização a cada minuto, atualização de localização com frequência incorreta e inconsistências nas próprias localizações. Nesse contexto é importante destacar que a coleta de dados dos veículos não é realizada de forma síncrona e contínua. Os dados enviados pelos veículos para o servidor central têm pelo menos um minuto de intervalo e não são sincronizados dentro desse minuto, impossibilitando uma comparação direta e instantânea de posições de veículos.

A Figura 4.1 mostra o número de atualizações no dia 07 de outubro de 2014, durante o período de 24 horas do ônibus C302810. O ônibus escolhido é um dos 3 que mais gerou atualizações nesse dia. Nessa figura é possível observar a quantidade próxima a 60 atualizações por hora durante todo o dia. Nas duas últimas barras, entre 23:00h e 23:59h, a quantidade de localizações enviadas é menor do que o restante do dia. Não existe nenhuma informação provida pela prefeitura que explique o motivo da perda de algumas localizações durante o dia ou durante esse horário específico.

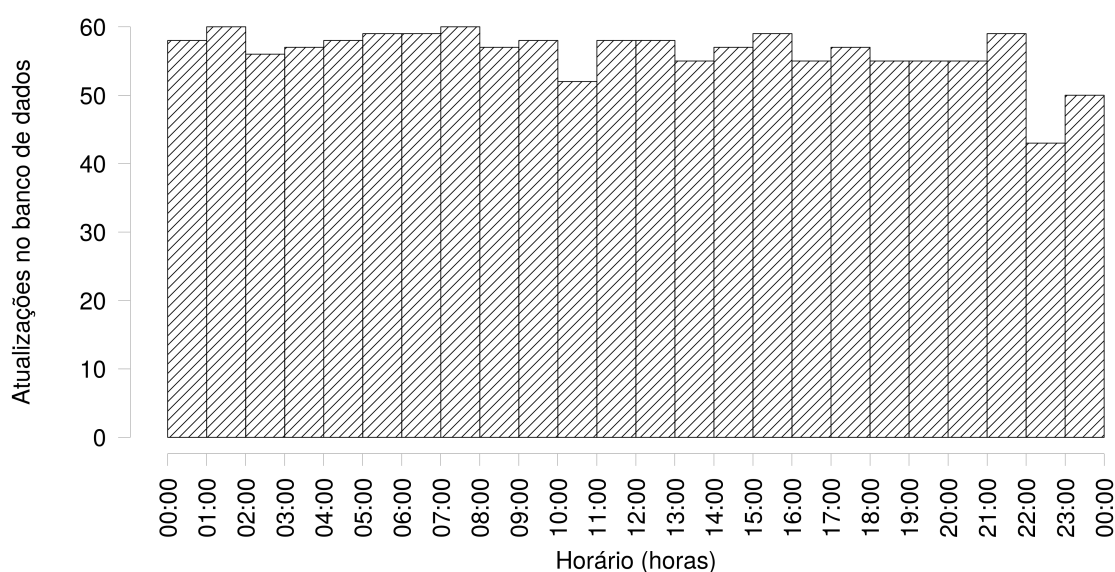


Figura 4.1: Número de atualizações do ônibus C302810 no banco de dados no dia 07/10/2014.

Como os dados não são sincronizados, não é possível comparar as localizações dos veículos de forma instantânea, ou seja, observar a localização de cada um dos veículos em um determinado instante (segundo) e calcular as possíveis interações entre eles. Duas abordagens distintas podem ser utilizadas para possibilitar a identificação de veículos próximos, que são: interpolar os dados entre cada localização obtida de cada veículo, ou ampliar o tempo utilizado para a comparação de localizações.

Interpolar os dados de localização significa, a partir de duas posições consecutivas de um veículo, inferir uma velocidade média e um trajeto para o mesmo e criar pontos de localização, sinteticamente, nesse intervalo. Essa abordagem leva diretamente ao aumento da quantidade de dados a serem analisados, já que serão criadas novas localizações dos veículos. No caso do banco de dados de ônibus, o menor intervalo entre as posições é de 1 minuto, o que leva inclusão de 59 localizações de veículos. Como 90% dos intervalos entre as atualizações de posições tem até em 7 minutos, em determinados casos é necessário criar 420 localizações que os veículos estiveram.

Além disso, ao tentar identificar uma velocidade média e um trajeto para os veículos, é inserida uma incerteza na localização reportada pelos veículos, pois os veículos não mantêm uma velocidade constante e nem sempre apresentam trajetos retilíneos. Como exemplo, um ônibus em sua velocidade máxima permitida (90 km/h) percorre 1.500 metros em um minuto. Esse é um longo intervalo de espaço, capaz de introduzir grandes incertezas se o trajeto não for corretamente calculado.

A segunda abordagem, ampliar o tempo utilizado para a identificação da localização, significa considerar um intervalo de tempo maior na comparação de posições, ou seja, criar uma janela de tempo em que a última localização conhecida dos veículos será utilizada para a comparação. Além da janela de tempo, para a análise, é necessário criar um raio de alcance para as comparações. Essa abordagem leva a restrições dos cálculos que podem ser executados com os dados, pois é necessário levar em consideração que os veículos se movem nesse intervalo de tempo.

Como ambas as abordagens introduzem erros no posicionamento dos veículos, foi escolhida a janela de tempo para a análise de viabilidade (análise preliminar) por ser de mais fácil implementação. Mesmo assim, existem vantagens ao usar uma janela de tempo. Essa abordagem não implica na criação de localizações fictícias para os veículos, o que mantém a mesma quantidade de dados a serem analisados e não leva a suposições quanto ao trajeto dos veículos, sendo utilizadas as localizações realmente reportadas pelos veículos.

Com o intuito de descobrir uma maneira factível para o processamento dos dados obtidos e levando em consideração que a análise dessa magnitude de dados coletados não é trivial, foi idealizada e realizada uma análise preliminar para identificar os cálculos necessários para localizar as interações que podem ocorrer entre veículos.

4.1.1 Análise preliminar

A primeira análise consiste em, a partir do banco de dados de todos os ônibus, localizar todos os ônibus que reportam a localização próxima à posição do ônibus atualmente analisado, em um intervalo de tempo definido. Dessa forma é possível

identificar quais ônibus estão próximos, ou seja, que estão no raio de alcance do ônibus. Os dados dos ônibus localizados devem ser armazenados para determinar em seguida as estatísticas de interesse.

Dessa forma, esta seção está dividida em escolha de parâmetros para a análise na Seção 4.1.1.1, onde são discutidos os valores utilizados na análise, a Seção 4.1.1.2 detalha como a análise por janelas de tempo é realizada e são apresentados os dados obtidos na Seção 4.1.1.3.

4.1.1.1 Escolha de parâmetros para a análise

Para a análise preliminar, foi escolhida a segunda abordagem, que consiste na comparação por intervalo de tempo ou janela de tempo. Ela foi escolhida por não ser necessária a introdução de novos dados no banco de dados e consequentemente não introduzindo erros de localização. Além disso, como menos dados de localização são utilizados, o custo das análises realizadas é menor.

A análise de todos os contatos de cada um dos veículos durante as 24 horas é proibitiva em função do tempo necessário para a sua execução, como será explorado na Seção 4.1.2 a partir dos dados da análise preliminar. Dessa forma, é analisado somente uma porção dos dados, utilizando somente o banco de dados dos ônibus e somente uma porção dos veículos.

Foi escolhido o banco de dados de ônibus pela sua regularidade em termos dos intervalos de atualizações, que tem os intervalos mais homogêneos do que o banco de dados de táxis. A regularidade faz com que os cálculos sejam mais fáceis de serem observados e que possam ser extrapolados para todos os veículos de forma mais fidedigna.

Além disso, a avaliação inicial é realizada somente com 10 ônibus distintos pertencentes a 10 linhas distintas. Esses ônibus foram escolhidos de forma aleatória, sem qualquer restrição. Os dados dos ônibus escolhidos estão exibidos na Tabela 4.1.

Tabela 4.1: Escolha dos ônibus analisados.

Ônibus	Linha	Quantidade de entradas
C516370	371	873
C825600	363	904
A481390	415	1157
B516440	624	905
A275950	433	945
C302230	550	946
D877380	2336	940
A291850	474	866
C475260	353	817
C302810	886	1350

Os parâmetros de análise são o tamanho da janela de tempo e o raio de busca de outros veículos. O primeiro estudo a ser realizado advém das diferentes possibilidades de variação destes parâmetros.

Primeiro parâmetro, a janela de tempo, influencia diretamente no tempo de execução, já que torna necessária a determinação da distância entre mais pontos. Para a escolha de parâmetros de análise foram escolhidas 4 janelas de tempo: 1 minuto, 2 minutos, 5 minutos e 10 minutos. Esses valores de variação foram escolhidos devido ao tempo entre localizações reportadas pelos veículos, que no melhor caso é de 1 minuto e em 90% dos casos tem intervalo de 7 minutos. O valor de 1 minuto representa o melhor caso, se todos os veículos mandassem a sua localização a cada minuto. O valor de 10 minutos é utilizado para garantir que a maior quantidade de ônibus possível seja analisado.

A partir do ponto que se quer analisar, a janela é dividida ao meio, buscando informações de horários anteriores e posteriores ao horário do ponto analisado. Dessa forma, se o horário analisado é 08:00h e a janela é de 2 minutos, as buscas no banco de dados serão dos pontos que estão entre 7:59h e 8:01h.

O segundo parâmetro que sofre variação é o raio de busca de outros veículos, ou seja, o raio de alcance. O alcance depende de vários fatores, como tecnologia utilizada, obstruções no ambiente, quantidade de interferência no local, quantidade de ônibus próximos, tempo de contato, velocidade de contato, entre outros fatores [19]. Esse fator não modifica o tempo de execução do algoritmo, já que não modifica a quantidade de pontos analisados de forma drástica. Para a escolha de parâmetros de análise foram escolhidos 4 raios de alcance: 50 metros, 100 metros, 200 metros e 300 metros. Esses alcances representam as tecnologias mais comumente utilizadas em redes veiculares [25].

4.1.1.2 Método de análise

O algoritmo de análise preliminar consiste em executar, para cada uma das posições reportadas no banco de dados dos 10 veículos selecionados, uma busca por todas as entradas no banco de dados que estão dentro da janela de tempo. Com a localização desses outros veículos, é determinado se a distância entre eles é maior ou menor do que um raio predeterminado. As entradas correspondentes são armazenadas para determinar as métricas desejadas.

Nesse algoritmo, a parte mais custosa é a localização dos veículos que estão dentro do raio proposto, pois é necessário calcular a distância a partir de coordenadas de GPS, levando em consideração o formato da terra. A quantidade de outros veículos que devem ser analisados contribui diretamente para o tempo de execução do algoritmo. A parte de análise das métricas após a descoberta dos veículos tem custo computacional baixo e não representativo em comparação com a descoberta

dos veículos.

A partir desse algoritmo foram obtidas as seguintes métricas:

- quantidade de contatos;
- média de contatos;
- média de tempo dos contatos;
- proporção de contatos seguidos;
- número de veículos contactados no ponto de contatos máximos;
- número de veículos únicos contactados;
- número de linhas únicas; e
- velocidade média.

De forma a simplificar a leitura e compreensão dos dados, evitando a apresentação repetitiva de tabelas, somente os dados relacionados quantidade de contatos, média de contatos e média de tempo dos contatos são exibidos a seguir. As demais tabelas são apresentadas no Apêndice A, porém suas características são avaliadas e comentadas neste capítulo.

4.1.1.3 Dados obtidos

Diversas métricas podem ser retiradas dos contatos do ônibus. Inicialmente, são analisados os dados de contato com outros ônibus, determinando quantos contatos são realizados e quais são as características desses contatos, como o número de ônibus distintos, o número de linhas distintas contactadas, qual a velocidade média de contato e qual a quantidade de contatos no ponto máximo de contatos.

4.1.1.3.1 Quantidade de contatos

A primeira métrica analisada é a quantidade de contatos que um ônibus teve durante um dia, que influencia diretamente a capacidade de troca de informações com outros veículos. Dentro da janela de tempo e do raio analisados, durante o período de 24 horas, os ônibus têm a quantidade de contatos descrita na Tabela 4.2.

Apesar das linhas analisadas serem de áreas distintas da cidade e serem controladas por empresas diferentes, as quantidades de contatos com outros ônibus são semelhantes, tendo diferenças menores do que o triplo do menor valor quando são considerados raios de 50 metros e de menos do que o dobro quando são considerados raios de 300 metros.

Tabela 4.2: Quantidade de contatos.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	3714	3252	3426	3023	2446	3010	2638	2062	3254	27976
100	1	7759	6055	6111	5487	4165	6010	4844	3406	8058	55303
200	1	12882	11773	12417	8710	8714	11986	9739	6284	18298	94722
300	1	16408	18035	19520	11542	14598	17731	16015	10534	23683	159138
50	2	6615	5710	5655	5152	3928	5071	4400	3286	5710	54636
100	2	14734	11328	10944	10160	7378	11030	8740	5975	15143	109087
200	2	25008	22843	23594	16667	16416	22932	18272	11765	35411	187496
300	2	32109	35325	38005	22568	28288	34385	30628	20214	46223	315793
50	5	15108	12771	12121	11331	8653	11511	9693	6952	12999	133852
100	5	35342	27050	25054	23752	17396	26328	20601	13550	36307	268729
200	5	61021	55447	56869	39931	39837	56155	44252	27866	86396	463039
300	5	78750	86335	93181	54833	69433	84505	75147	49092	112931	780686
50	10	28966	23566	22199	21305	15923	21608	18229	12733	24772	264660
100	10	69039	51850	47508	45801	33421	50503	39810	25734	70780	532604
200	10	120334	108271	110381	77919	77833	109590	87416	53979	169183	918174
300	10	155603	169426	182043	107600	136615	165286	148794	95974	221810	1549602

Outro dado coletado é a quantidade de ônibus únicos que são contatados em 24 horas, que representa a quantidade de veículos distintos contatados, apresentado na Tabela A.2. A quantidade de ônibus únicos cresce de acordo com o raio e com a janela de tempo. A variação entre o raio de 50 metros e a janela de 1 minuto e o raio de 300 metros e a janela de 10 minutos pode chegar a 8 vezes, no caso do ônibus D877380. A quantidade de ônibus únicos contatados influencia diretamente na disseminação dos dados. Um ônibus que se conecta a vários ônibus distintos, dependendo do protocolo de disseminação utilizado, tem uma maior possibilidade de disseminar uma mensagem para outros veículos necessitando de poucas retransmissões da mensagem.

O número de linhas distintas contatadas relaciona-se a possibilidades de enviar dados para trajetórias distintas da trajetória do ônibus analisado. Este dado é apresentado da Tabela A.6. Quanto maior o número de contatos com linhas diferentes, maior o alcance dos dados em relação à distância. Dessa forma, um ônibus que tem contatos com uma grande quantidade de linhas, é capaz de espalhar os dados de forma mais eficaz entre as diversas regiões da cidade, precisando de menos intermediários para que o dado chegue ao destino. O número de linhas únicas aumenta linearmente de acordo com o raio de contato entre os ônibus. A variação de linhas contatadas é pequena entre uma janela de 1 minuto e uma janela de 10 minutos, indicando que contatar linhas distintas tem uma maior relação com o alcance do que com a quantidade de ônibus contatados, representado aqui pela variação da janela.

4.1.1.3.2 Média de contatos

A Tabela 4.3 mostra o número médio de contatos que ocorre em cada janela analisada. Esse dado indica a quantidade de conexões possíveis com outros ônibus

em uma determinada janela de tempo. Desse número é possível identificar que ao aumentar a janela e o raio de busca, o número de ônibus contatados é ampliado, em consequência, o número de transmissões simultâneas também é ampliado, podendo levar a uma maior interferência nas transmissões de dados. A quantidade média de contatos é bastante distinta entre os ônibus, porém o número médio de contatos sempre aumenta de acordo com o raio e de acordo com a janela de tempo, mostrando a importância da definição correta do equipamento de radiocomunicação. Através da escolha da tecnologia pode-se reduzir as interferências, porém também o número de vizinhos de um nó.

Tabela 4.3: Média de contatos por ponto.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	4,3	3,6	3,0	3,3	2,6	3,2	2,8	2,4	4,0	20,7
100	1	8,9	6,7	5,3	6,1	4,4	6,4	5,2	3,9	9,9	41,0
200	1	14,8	13,0	10,7	9,6	9,2	12,7	10,4	7,3	22,4	70,2
300	1	18,8	20,0	16,9	12,8	15,5	18,7	17,0	12,2	29,0	117,9
50	2	7,6	6,3	4,9	5,7	4,2	5,4	4,7	3,8	7,0	40,5
100	2	16,9	12,5	9,5	11,2	7,8	11,7	9,3	6,9	18,5	80,8
200	2	28,7	25,3	20,4	18,4	17,4	24,2	19,4	13,6	43,3	138,9
300	2	36,8	39,1	32,9	24,9	29,9	36,4	32,6	23,3	56,6	233,9
50	5	17,3	14,1	10,5	12,5	9,2	12,2	10,3	8,0	15,9	99,2
100	5	40,5	29,9	21,7	26,3	18,4	27,8	21,9	15,7	44,4	199,1
200	5	69,9	61,3	49,2	44,1	42,2	59,4	47,1	32,2	105,8	343,0
300	5	90,2	95,5	80,5	60,6	73,5	89,3	79,9	56,7	138,2	578,3
50	10	33,2	26,1	19,2	23,5	16,9	22,8	19,4	14,7	30,3	196,0
100	10	79,1	57,4	41,1	50,6	35,4	53,4	42,4	29,7	86,6	394,5
200	10	137,8	119,8	95,4	86,1	82,4	115,9	93,0	62,3	207,1	680,1
300	10	178,2	187,4	157,3	118,9	144,6	174,7	158,3	110,8	271,5	1147,9

A quantidade de veículos no ponto onde a maior quantidade de veículos são contatados é analisada e é apresentada na Tabela A.5. Há um aumento linear na quantidade de ônibus contatados no instante com maior número de contatos simultâneos, em relação à janela de tempo. A variação do raio aumenta linearmente com o raio até 200 metros, porém, entre 200 e 300 metros o número de contatos máximos simultâneos é pouco alterado. No caso do ônibus C516370, o número de contatos é o mesmo para 200 e 300 metros. Como o número de ônibus contatados aumenta com o alcance, a interferência é maior para alcances maiores. Entretanto, o número máximo de contatos tem uma menor alteração entre 200 e 300 metros ao comparar com outros raios para alguns ônibus, o que indica que nem sempre usar tecnologias com maior alcance não traz necessariamente maior interferência.

4.1.1.3.3 Características dos contatos entre ônibus

A Tabela 4.4 apresenta dados dos tempos médios de contato entre o ônibus analisado e os outros ônibus localizados. O tempo, representado na tabela em segundos,

apresenta o valor máximo na maioria dos casos entre 50 e 100 metros de raio. Ao aumentar a janela, mais ônibus que não estão nas mesmas vias e conseqüentemente não fazem caminhos semelhantes, são contatados. Esses ônibus naturalmente têm uma menor oportunidade de contato, reduzindo a média de tempo de contato. A janela de contato, de forma semelhante, também amplia o número de ônibus contatados que não têm itinerário parecido com o do ônibus que está sendo analisado, fazendo com que o tempo médio de contato decresça na maioria dos casos.

Tabela 4.4: Tempo médio dos contatos (segundos).

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	2555	2654	3254	2520	2991	5313	2414	3630	1059	4609
100	1	3605	3852	4539	4955	4165	5262	4482	4556	1901	5572
200	1	3720	3876	5467	4859	4184	5453	4765	4180	2467	4723
300	1	3929	3355	4622	4477	4454	2576	4155	3683	2226	4182
50	2	2315	2400	2765	1780	1939	3172	2659	1930	1296	2905
100	2	3261	2416	2925	2312	2570	2374	3153	2138	1094	2959
200	2	2136	1623	2064	1858	1857	1166	2228	1669	1283	2025
300	2	1533	1067	1347	1363	982	701	1749	1154	953	2696
50	5	3001	2474	2277	3071	1816	2584	3453	2240	1245	2812
100	5	2508	1761	1941	2323	1766	1410	2795	2251	1077	2340
200	5	1222	1160	1206	987	989	976	1813	1204	744	1795
300	5	1058	713	676	785	591	606	1114	731	637	2297
50	10	3508	2753	2733	3428	2394	2280	3803	2463	1166	3262
100	10	2375	1625	1626	1793	1582	1187	2572	1749	1021	2156
200	10	963	926	824	821	646	649	1187	1042	709	2079
300	10	755	611	487	563	294	625	688	627	607	3048

Os contatos com outros ônibus são divididos em dois tipos: os contatos únicos são aqueles em que, para o mesmo ônibus em período correspondente a uma janela de tempo, o ônibus foi contatado somente uma vez. Os contatos múltiplos são aqueles em que o mesmo ônibus é contatado mais de uma vez na janela de tempo. Um contato múltiplo garante que o contato entre os dois ônibus é realizado por mais tempo, possibilitando uma maior troca de dados. Ao incrementar a janela e o raio, a proporção de contatos longos é maior, aumentando a oportunidade de cada ônibus contatado trocar dados de forma satisfatória.

O banco de dados apresenta informação de velocidade instantânea dos ônibus a partir dos dados do GPS. Para fins de análise, essa informação é utilizada como velocidade na hora do contato e é apresentada na Tabela A.10. A velocidade média de contatos é inferior a 15 km/h para todos os ônibus analisados, chegando a menos de 1 km/h em alguns casos (C302810), não tendo um padrão predefinido. Como os ônibus compartilham a mesma infraestrutura rodoviária, as condições de tráfego na via estão ligadas diretamente à velocidade dos ônibus. Além disso, como o raio de contato é pequeno, ônibus próximos tendem a ter a mesma velocidade. Aliado às condições de tráfego na via, como os intervalos de medições são longos, no melhor caso de 1 minuto, a probabilidade que dois ônibus em alta velocidade, ou mesmo de

um parado e de outro em alta velocidade, se encontrem é menor do que dois ônibus que estejam em velocidade baixa. Dessa forma, a velocidade média global é baixa.

4.1.2 Tempo de execução da análise preliminar

O algoritmo de análise preliminar foi implementado com a linguagem de programação R. Como dito, o tempo de análise depende diretamente dos parâmetros de análise escolhidos e do tamanho dos bancos de dados analisados, além do desempenho computacional. Nesse caso, as análises descritas foram executadas de forma simultânea em um núcleo de um servidor. O computador utilizado para as análises é um servidor com dois processadores Intel Xeon X5570, com frequência de 2,93 GHz e possuindo 96 GB de memória RAM. Três análises foram executadas de forma exclusiva, para determinar o menor tempo de execução possível para o algoritmo. A Tabela 4.5 na terceira coluna (Tempo de execução) exibe os tempos necessários para a execução do algoritmo de busca de todos os contatos de um ônibus, para os 10 ônibus selecionados, durante um dia. A quarta coluna (Tempo Total) exibe o tempo em dias necessário para a execução da análise para todos os veículos, onde o tempo total foi calculado ao extrapolar o tempo de execução em horas para todos os 12.456 ônibus, ou seja, o tempo de execução multiplicado por 1.245,6.

Tabela 4.5: Tempo de execução da análise preliminar.

Raio (metros)	Janela de tempo (minutos)	Tempo de execução (horas)	Tempo total (dias)	Observações
50	1	8,15	422,99	
100	1	8,18	424,54	
200	1	8,07	418,83	
300	1	4,99	258,98	Exec. exclusivamente
50	2	15,47	802,89	
100	2	14,36	745,28	
200	2	15,91	825,73	
300	2	9,75	506,03	Exec. exclusivamente
50	5	32,02	1661,84	
100	5	33,78	1753,18	
200	5	33,05	1715,30	
300	5	23,37	1212,90	Exec. exclusivamente
50	10	56,18	2915,74	
100	10	57,58	2988,40	
200	10	57,82	3000,86	
300	10	60,46	3137,87	

Para um raio de 300 metros, com uma janela de 5 minutos, a análise preliminar leva 23,37 horas para ser executada para 10 ônibus. Ao extrapolar esse dado para

todos os ônibus do banco de dados, no caso 12.456, seriam necessários mais de 1.200 dias para a computação completa dos contatos de todos os ônibus. O algoritmo proposto foi implementado sem paralelismo, porém o mesmo é paralelizável. Se o algoritmo fosse implementado de forma paralela, ainda que dividido em múltiplas partes e distribuído a diversos computadores, seria necessária uma quantidade significativa de processamento paralelo para reduzir os 1.200 dias estimados para um tempo satisfatório. Supondo que, de forma ideal, cada divisão do algoritmo seja capaz de diminuir o tempo de processamento pela metade, seriam necessárias 1.200 divisões, ou 1.200 núcleos de computadores, para reduzir os 1.200 dias para 1 dia de processamento.

4.1.3 Resultados da análise preliminar

Para ser possível identificar quais regiões da cidade estão interligadas e quanta informação é possível transmitir na cidade, é necessário que se use os dados da maior quantidade possível dos veículos que nela circulam, no caso estudado, ônibus e táxis. A partir da análise preliminar, foi identificado que não é viável, sem um grande esforço computacional, a análise de todos os contatos possíveis entre veículos.

No caso do banco de dados dos ônibus, para procurar os ônibus vizinhos de um ponto em um determinado horário, é necessário buscar qual é a distância informada entre o ponto de referência e os outros ônibus. Para o banco de dados de ônibus, em média, são necessárias 5.000 comparações de distância a cada intervalo de um minuto para o qual é realizada uma busca, ou seja, para cada localização de cada ônibus são necessárias 5.000 medições de distância de coordenadas de GPS. Para analisar 24 horas de tráfego, precisa-se analisar 1.440 minutos, ou seja, 7,2 milhões de vezes por dia por ônibus. Ao analisar o período de 24 horas em um banco de dados de 12.456 ônibus, são necessárias até 450 bilhões de comparações de distância entre os ônibus. Dessa forma, mostra-se necessário o uso de ferramentas de análise de grandes massas de dados para possibilitar responder a essas perguntas.

A partir dos dados coletados e da análise preliminar é possível identificar que a janela de tempo e o raio de alcance são fundamentais para a determinação dos contatos entre os veículos e do tempo de processamento.

4.2 Ferramentas de análise

Conforme indicado, o custo computacional para analisar todos os possíveis contatos de veículos é potencialmente muito alto. Para este fim, é necessário calcular, para cada veículo e em todos os momentos, os veículos que estão na sua vizinhança conforme definida pela tecnologia de comunicação sem-fio utilizada. Nesse caso,

como o banco de dados é de tamanho significativo, o processo de determinação dos contatos é caro e demorado. Além disso, para determinar a distância entre dois pontos usando coordenadas GPS, é necessário considerar a curvatura da Terra para evitar erros de distância, sendo que estes cálculos envolvem aritmética de ponto flutuante.

Existem diversas formas de analisar uma grande quantidade de dados espaciais dependentes do tempo aplicadas ao caso estudado [60–63], que vão de algoritmos para determinar a distância entre todos os pontos até algoritmos adaptados a tratar grandes dados, que amostram um conjunto de pontos ou reduzem a área analisada ou ainda agrupam os dados espaciais (clusterização).

Um dos objetivos deste trabalho é investigar a viabilidade de uma determinada tecnologia sem-fio, considerando o raio de alcance nominal. Para este fim, adotar uma técnica de agrupamento de dados espaciais se adapta de forma adequada ao problema. De forma mais específica, considera-se um algoritmo de clusterização baseado em grade, ou *grid-based clustering*, para agrupar as localizações reportadas pelos veículos. Estudos realizados por Kolatch [64] sugerem o algoritmo de clusterização baseado em grade STING (*STatistical INformation Grid-base method*) como sendo o algoritmo de menor complexidade para este fim.

Dessa forma, esta seção apresenta a ferramenta de clusterização STING, uma dentre as várias possibilidades para a análise da grande quantidade de dados de localização dos veículos, fundamentando a escolha realizada nesta tese.

4.2.1 Algoritmo de clusterização STING

Para determinar o número de contatos entre veículos, é necessário, para cada veículo, procurar todos os demais veículos que estão próximos. Se nenhuma restrição for usada, cada veículo deve determinar a posição de todos os outros veículos para determinar os contatos, trazendo complexidade para $O(n^2)$ operações, onde n é o número de localizações de veículos no banco de dados. Neste caso, uma vez que o banco de dados contém uma grande quantidade de dados e a determinação da distância através das coordenadas não é trivial, o processo de determinação dos contatos é dispendioso e demorado.

O algoritmo STING [65] consiste em dividir as regiões de análise no menor tamanho desejável, usando regiões de mesmo tamanho e forma para criar uma grade com células fixas. As células de menor tamanho possível formam a camada de nível mais baixo. Para cada célula desta grade é calculado um parâmetro de relevância, no caso estudado, o número de localizações reportadas por veículos é o parâmetro de relevância escolhido. Após a criação da camada de nível mais baixo, novas camadas são formadas mesclando em uma célula de tamanho superior várias células

da camada inferior. É criada dessa forma uma estrutura hierárquica, composta por diversas camadas.

A criação dessa estrutura objetiva a redução da complexidade em termos do número de operações. Para realizar uma busca nas células, a pesquisa deve começar na camada superior, composta pelo agrupamento das diversas camadas inferiores, que seleciona as células que correspondem a uma especificação desejada. Em seguida, a busca continua somente no nível imediatamente inferior das células selecionadas anteriormente, descendo os diversos níveis até que as especificações desejadas sejam atingidas.

O algoritmo STING tem na fase de montagem da grade complexidade da ordem de $O(n)$, onde n é o número de localizações de veículos no banco de dados. Para as demais operações que envolvem as células, a complexidade torna-se da ordem de grandeza das células, ou seja, a dimensão k da grade, ou $O(k)$. Uma das principais vantagens desse método é a baixa complexidade, levando a uma operação mais rápida em comparação com outros algoritmos de clusterização, principalmente no uso de grandes dados. A complexidade desse método em comparação com outros métodos de clusterização é apresentada por Kolatch em [64].

No entanto, o método de clusterização baseado em grade tem desvantagens que, se não forem devidamente tratadas, podem levar a resultados ruins. A definição de regiões pode não ser ótima, o que significa que a escolha de limites das células e da localização das mesmas é crucial. Na verdade, ao escolher a localização das células no início da execução do algoritmo STING, algumas regiões podem não ser definidas exatamente em locais de aproveitamento máximo, ou locais com o maior número de veículos. Para minimizar esse problema, é necessário ajustar o tamanho das células para que não sejam muito pequenas (levando a um alto tempo de processamento) ou muito grandes (levando a perda de resolução para análise de dados), o que afeta diretamente o desempenho do algoritmo.

A Figura 4.2 mostra a aplicação do método STING no conjunto de dados de localização de veículos da cidade do Rio de Janeiro, descrito na Capítulo 3. Como cada camada corresponde a um tamanho de célula diferente, para analisar uma determinada tecnologia é necessário escolher uma camada específica, neste caso, uma célula quadrada com lado de 300 metros. O valor de cada célula corresponde ao número de todos os registros de localização GPS de todos os veículos dentro da célula durante as 24 horas. As 100 células com mais registros de localização (nós os chamamos de clusters) são destacadas na Figura 4.2. Nesse caso, cada um dos maiores clusters contém da ordem de 10.000 localizações de veículos.



Figura 4.2: Mapa do Rio de Janeiro exibindo os 100 maiores clusters de 300 m.

4.3 Métodos de transmissão de dados

A partir da escolha do método de clusterização por grades, ou *grid-based clustering*, é preciso definir quais são os possíveis cenários para a transmissão de dados. Como o algoritmo necessita da escolha dos parâmetros de tamanhos da grade, ou tamanho dos clusters, é essencial identificar possíveis valores para essa variável.

Os cenários de transmissão de dados possíveis utilizando os veículos de transporte urbano são diversos, porém algumas tecnologias de transmissão sem-fio são mais mencionadas na literatura, como por exemplo a tecnologia bluetooth (IEEE 802.15.1), a IEEE 802.11, nas suas diversas variantes, principalmente a IEEE 802.11p, IEEE 802.11g e IEEE 802.11n, além de redes de telefonia celular. Baseando-se nas características atuais de alcance das tecnologias de transmissão de dados sem-fio, é possível agrupar os seus níveis de alcance.

4.3.1 Tecnologias de transmissão

Redes sem-fio podem ser formadas em locais inesperados de maneiras inesperadas, principalmente devido à sua característica de ubiquidade. Por isso, Dhillon, Huang e Viswanathan em [66] publicaram um estudo onde caracterizam a comunicação de dados de máquinas para máquinas (M2M) em duas dimensões: alcance e mobilidade. O alcance refere-se à disseminação geográfica dos dispositivos enquanto a mobilidade está relacionada a se o dispositivo se move e se essa comunicação deve ser realizada enquanto o dispositivo está estático ou em movimento.

Assim sendo, as aplicações variam enormemente, desde dispositivos conectados

no mesmo ambiente, sem mobilidade, a dispositivos em um ambiente amplo, percorrendo vários quilômetros por dia. Essa variedade significa que nenhuma tecnologia única e nenhum dispositivo único é capaz de ser otimizado para uso em todas as aplicações. Para uma melhor ilustração, a Tabela 4.6, proposta por Dhillon *et al.* [66], apresenta 5 categorias diferentes para aplicações M2M.

Tabela 4.6: Visão geral das aplicações M2M.

Aplicação	Alcance	Mobilidade	Redes possíveis
<ul style="list-style-type: none"> • Carro conectado • Controle de frotas • Monitoramento remoto de saúde 	~1000m	Sim	<ul style="list-style-type: none"> • Celular • Satélite
<ul style="list-style-type: none"> • Medidor de consumo inteligente • Monitoramento de vagas 	~1000m	Não	<ul style="list-style-type: none"> • Celular • Dedicado
<ul style="list-style-type: none"> • Rastreamento de patrimônio • Logística de armazéns 	~100m	Sim	<ul style="list-style-type: none"> • WiFi • RFID
<ul style="list-style-type: none"> • Automação industrial • Automação residencial 	~10m	Não	<ul style="list-style-type: none"> • Zwave • Zigbee • WiFi • Powerline
<ul style="list-style-type: none"> • Atividade pessoal • Rastreamento de objetos pessoais • Pontos de venda 	~1m	Não	<ul style="list-style-type: none"> • Bluetooth • NFC

Fonte: Adaptado de Dhillon *et al.* [66] .

Dhillon apresenta as possíveis aplicações que abrangem toda uma cidade, direcionando a proposta da rede modelada. Desta forma, este trabalho não visa uma aplicação definitiva, no entanto, é direcionado para um perfil específico de comunicação de dados: ambiente de comunicações de longo alcance (mais de 1000m) e com ou sem mobilidade, ou as duas primeiras linhas da Tabela 4.6. Esses cenários geralmente são baseados em redes celulares, como tecnologias 4G e 5G, e redes dedicadas, implementadas exclusivamente à aplicação [67]. Além das tecnologias listadas por Dhillon, é possível utilizar o IEEE 802.11p aplicada diretamente á rede veicular.

4.3.2 Rede modelada

A rede modelada é composta por dois atores, por todos os veículos (ônibus e táxis) com equipamentos de rádio e por pontos de acesso localizados nos clusters mais movimentados, ambos utilizando uma tecnologia de transmissão escolhida. Além disso é idealizado que todos os pontos de acesso apresentam um enlace ou outro método de transmissão para a Internet, ou uma rede privada. Como os pontos de acesso têm conexão com uma rede maior, não é necessária a retransmissão de mensagens na rede proposta. Outras propostas de rede, com menos enlaces ligando

os pontos de acesso e com retransmissões entre os pontos de acesso é possível, porém não é estudada nesta tese.

Transmissões entre dispositivos espalhados pela cidade e veículos fora da região dos clusters não são consideradas nesta análise. Por exemplo, no caso de uma aplicação de Internet das Coisas (IoT – *Internet of Things*) contendo sensoriamento remoto, os dados trafegados geralmente são da ordem de bits. Esses dados são coletados no caminho dos veículos ou gerados nos próprios veículos e não necessariamente estão localizados nos clusters utilizados.

Para inferir a transmissão sem precisar selecionar uma única tecnologia sem-fio, escolhem-se múltiplos tamanhos de grades de acordo com as tecnologias disponíveis. A Tabela 4.7 mostra os tamanhos de grades simulados e as tecnologias mais comuns que podem ser empregadas em transmissões de dados para os veículos. A última coluna mostra a capacidade de largura de banda esperada de cada tecnologia.

Conforme apresentado na Tabela 4.7, a largura de banda associada às diferentes tecnologias varia enormemente. Para representar melhor essa diversidade, além do alcance variável, a largura de banda das análises não foi definida: em vez disso, utilizou-se um fator de multiplicação α (alfa) para modelar a capacidade dos enlaces sem-fio, enquanto o alcance é considerado em metros.

Tabela 4.7: Tecnologias de rede sem-fio consideradas.

Tecnologias	Alcance (metros)	Banda
<ul style="list-style-type: none"> • Bluetooth • Zigbee • Zwave 	50	~Mbps
<ul style="list-style-type: none"> • 802.11ac 	100	~Gbps
<ul style="list-style-type: none"> • 802.11n 	300	~100 Mbps
<ul style="list-style-type: none"> • 802.16 • 802.11p 	500	~100 Mbps
<ul style="list-style-type: none"> • LTE-A • Lora 	1000	~Gbps ~kbps

A variável alfa é medida em bits por segundo, ou bps, ou seja, para se obter a quantidade de dados trafegados, é necessário multiplicar alfa por um período de tempo, em segundos. Medindo o tempo em que os veículos permanecem no alcance de rádio do ponto de acesso de um cluster, pode-se estimar a capacidade de transmissão de dados e os atrasos envolvidos na transmissão de dados através da rede veicular.

Assume-se que o número de pontos de coleta (pontos de acesso) espalhados pela cidade é limitado. Portanto, para determinar os melhores pontos da cidade para instalá-los, utiliza-se o algoritmo de clusterização. A metodologia proposta seleciona

os locais que são mais visitados pelo maior número de veículos como locais propícios para a instalação de um ponto de coleta, ou seja, o número de clusters analisados pelo algoritmo é finito e determinado de forma empírica.

Durante as próximas análises são sempre selecionadas as 100 maiores células, que passa-se a chamar simplesmente de clusters, como sendo o objeto da análise. Chegou-se a esse número de forma empírica, a partir da quantidade de veículos analisados (será apresentada na seção 4.4.3), tentando se aproximar do uso de 100% dos veículos para a maioria dos casos, além de tentar balancear o tempo de execução do algoritmo, o número de atualizações em cada uma das células (será analisado no Capítulo 5) e um valor factível para a implementação de pontos de acesso com conexão a Internet na cidade do Rio de Janeiro. É importante ressaltar que o número de clusters utilizados está diretamente ligado à capacidade da rede, em que uma rede com um número maior de clusters tem uma capacidade maior por ter mais pontos de acesso.

Finalmente, ao combinar as medidas de capacidade obtidas com os requisitos de aplicações, pode-se apontar quais aplicações determinada rede veicular é capaz de suportar. Ao conhecer os requisitos do aplicativo, como o atraso aceitável e a quantidade de dados coletados dos sensores, é possível determinar a tecnologia sem-fio e a quantidade de clusters contendo pontos de acesso que são necessários para suportar as aplicações.

É relevante destacar que o alcance das tecnologias representa o raio de alcance por meio da projeção de um círculo ao redor do equipamento, o que é diferente da grade escolhida no algoritmo de clusterização. A grade do algoritmo é quadrada e o lado do quadrado é utilizado como delimitador. Portanto, a distância do centro do quadrado até o meio de sua lateral é a metade do lado e até os vértices do quadrado é metade da diagonal do quadrado. Essa abordagem escolhida é conservadora para comparar o alcance das tecnologias já que a projeção do alcance das tecnologias sem-fio, que normalmente são em forma de círculo, utilizando o raio de alcance, e a projeção utilizada, que é composta por um quadrado com o lado do tamanho do alcance, são distintas. O quadrado utilizado tem a área 3 vezes menor do que a área do círculo de projeção do alcance.

4.4 Cálculo de deslocamento dos veículos

A partir das metodologias destacadas neste capítulo foi elaborado um algoritmo para identificar e contabilizar os deslocamentos realizados pelos veículos, o que inclui identificar uma viagem entre clusters, contabilizar o número e o tempo dessas viagens e o tempo que os veículos permanecem em cada cluster.

4.4.1 Algoritmo de cálculo de viagens

O algoritmo é desenvolvido com base no método de clusterização por grade, que agrupa veículos em cada uma das células. Em seguida, analisa-se as 100 maiores células, que chamamos de clusters, em termos de número de veículos que reportaram a sua localização naquela célula. As seguintes características são então calculadas: número de viagens entre pares de clusters; tempo de viagem entre pares de clusters; e tempo de permanência em cada cluster. Uma viagem entre os clusters A e B é definida como a existência de um veículo que relatou sua posição dentro do cluster A seguido por relatar a sua localização dentro do cluster B.

O Algoritmo 1 é utilizado para analisar as características das viagens. As entradas do algoritmo são a lista de veículos $a[]$ em cada cluster e o número de clusters analisados N (linha 1 do Algoritmo 1). As variáveis $line$ e $column$ têm o objetivo de indexar a variável de cluster para percorrer todos os clusters (linhas 2 e 3). Quando o cluster de origem é o mesmo que o de destino, ou seja, $line = column$, o algoritmo não executa nada (linha 4). A função `intersection(a[line],a[column])` recebe como entrada o conjunto de veículos em dois clusters diferentes e retorna b , o subconjunto de veículos detectados em ambos os clusters (linha 5). A função `vehiclesTimeLocation()` recebe uma lista de todos os veículos que são detectados nos clusters $line$ e $column$ e retorna uma lista com o tempo reportado por cada veículo nos clusters $timelineLine$ e $timelineColumn$ (linhas 7 e 8). Para todos os veículos detectados em cada cluster (linhas 11 a 30), o algoritmo armazena o número total de viagens `travels[line,column]`, o tempo total de viagem `travelTime[line,column]`, o tempo de permanência total no cluster i , `stayTimeI[line,column]` e o tempo de permanência total no cluster j , `stayTimeJ[line,column]`.

A Figura 4.3 ilustra as diferentes medidas de tempo coletadas utilizadas para o cálculo das viagens. Cada coluna representa a linha do tempo de cada veículo em um cluster específico, ou seja, os horários que determinado veículo reportou a sua localização em cada um dos clusters. O veículo 1 está no cluster A às 00:01h, 00:02h, 00:03h, 01:00h e 02:00h. O mesmo veículo relata sua localização no cluster B às 00:33h, 00:34h e 02:30h. O tempo de viagem é definido pela última vez que o veículo reportou estar no cluster A até a primeira vez que o veículo reportou estar no cluster B; entretanto, somente após o horário que ele saiu do cluster A (linhas 12 e 18). Cada vez que uma viagem é identificada ela é contabilizada (linhas 16 a 18 e 25 a 27). Na Figura 4.3, o primeiro tempo de viagem do veículo 1 de A até B é de 30 minutos, a diferença entre 00:03h e 00:33h. O tempo de permanência no cluster A é calculado como o tempo que o veículo informa que está dentro do cluster A. Para o cluster A, é o tempo entre 00:01h e 00:03h, para o cluster B, o tempo entre

Algoritmo 1 Algoritmo de cálculo de viagens.

```
1: procedure VIAGENS( $a[]$ ,  $N$ ) ▷ Onde a-Matriz com os veículos nos clusters,  
   N-Número de clusters  
2:   for  $line \leftarrow 1, N$  do ▷ Linhas dos Clusters  
3:     for  $column \leftarrow 1, N$  do ▷ Colunas dos Clusters  
4:       if ( $line \neq column$ ) then ▷ Não executa se origem=destino  
5:          $intersect \leftarrow intersection(a[line], a[column])$   
6:         for all  $intersect$  do ▷ Analisando todos os veículos nas linhas e colunas  
7:            $timelineLine \leftarrow vehiclesTimeLocation(b[line])$   
8:            $timelineColumn \leftarrow vehiclesTimelocation(b[column])$   
9:            $timelineLinePointer \leftarrow 0$   
10:           $timelineColumnPointer \leftarrow 0$   
11:          for all  $timelineLine$  AND  $timelineColumn$  do  
12:            if ( $interval(timelineLine[tlLineIdx] > timelineColumn[tlColumnIdx])$ )  
   then  
13:             if  $interval < 5 * 60$  then  
14:                $stayTimeI \leftarrow stayTimeI + interval$   
15:             end if  
16:              $travelTime \leftarrow -interval$   
17:              $Travel \leftarrow TRUE$   
18:              $tlLineIdx \leftarrow tlLineIdx + 1$   
19:           else  
20:             if  $interval < 5 * 60$  then  
21:                $stayTimeJ \leftarrow stayTimeJ + interval$   
22:             end if  
23:              $tlColumnIdx \leftarrow tlColumnIdx + 1$   
24:             if  $Travel$  then  
25:                $Travel \leftarrow FALSE$   
26:                $travelsIJ \leftarrow travelsIJ + 1$   
27:                $totalTravelTime \leftarrow totalTravelTime + travelTime$   
28:             end if  
29:           end if  
30:         end for  
31:       end for  
32:     end if  
33:   end for  
34: end for  
35:    $travels[line, column] \leftarrow travelsIJ$  ▷ Número de viagens  
36:    $travelsTime[line, column] \leftarrow totalTravelTime$  ▷ Tempo total de viagens  
37:    $stayTimeI[line, column] \leftarrow stayTimeLine$  ▷ Tempo de permanência em i  
38:    $stayTimeJ[line, column] \leftarrow stayTimeColumn$  ▷ Tempo de permanência em j  
39: end procedure
```

00:33h e 00:34h. Neste exemplo, assume-se que não é possível avaliar se o veículo estava no cluster A entre 01:00h e 02:00h, então esse período não é contabilizado como tempo de permanência.

O tempo de permanência em um cluster é calculado pelo Algoritmo 1 do momento inicial em que um veículo relata sua localização dentro de um cluster (linhas 13 a 18) até o último relatório no mesmo cluster (linhas 10 a 27). Este tempo de

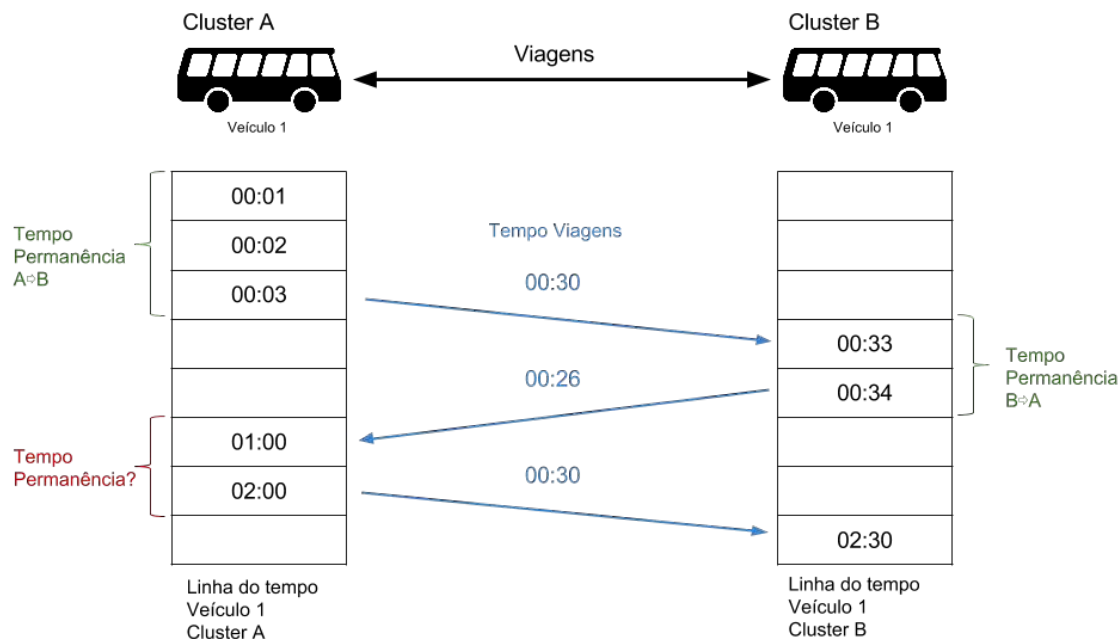


Figura 4.3: Representação de viagens entre dois clusters.

permanência é medido apenas para a mesma viagem, como no exemplo anterior. Para prevenir erros devido à perda de dados de localização de veículos, todos os dados de localização em um mesmo “tempo de permanência” são verificados. Apenas localizações reportadas consecutivamente, com intervalo de tempo inferior a 5 minutos, são contabilizadas (linhas 13 a 15 e 20 a 22). O intervalo de 5 minutos foi escolhido pois as informações de localização são esperadas uma vez por minuto para cada veículo. Escolher não considerar intervalos maiores do que 5 minutos significa que são necessárias 6 falhas de comunicação ou erros de banco de dados seguidas para que um intervalo seja considerado inválido. O intervalo de 5 minutos foi escolhido porque 85% dos dados de localização têm intervalos de atualização inferiores a 5 minutos no banco de dados coletado. -

Dessa forma, ao analisar o tempo que um veículo permanece no alcance de uma área determinada e o padrão do movimento do veículo, é possível inferir a capacidade da rede veicular dado o valor de α e o raio da tecnologia escolhida.

4.4.2 Tempo de execução do algoritmo de cálculo de viagens

O algoritmo desenvolvido foi, de forma semelhante à análise preliminar, implementado com a linguagem de programação R. O Algoritmo 1, por necessitar de uma menor quantidade de processamento, foi executado em um computador pessoal, com processador Intel Core i7 4790K e 3,80 GHz de frequência e possuindo 16 GB de memória RAM.

A Tabela 4.8 apresenta a quantidade de minutos necessária para a execução do

algoritmo. Cada linha representa um tamanho de célula e cada coluna representa o banco de dados utilizado na análise. Na coluna Ônibus são usadas como banco de dados somente as localizações de GPS dos ônibus, na coluna Táxis são utilizados somente os dados de localizações de GPS dos táxis e na última coluna os dados são combinados e analisados de forma simultânea.

Tabela 4.8: Tempo de execução do Algoritmo 1.

Alcance (metros)	Tempo de execução (minutos)		
	Ônibus	Táxis	Ônibus e Táxis
50	31	15	51
100	37	16	64
300	110	68	400
500	202	146	641
1000	304	300	875

A análise preliminar era executada somente em 10 ônibus pré-selecionados, já o Algoritmo 1 analisa o banco de dados de veículos em sua totalidade. A análise anterior, em seu pior caso, foi realizada em 60 horas, já o algoritmo atual, no caso similar é executado em menos de 2 horas, e, em seu pior caso, é executado em 14 horas em um computador de capacidade de processamento inferior. Essa métrica mostra que o algoritmo desenvolvido é menos custoso do que a procura de todos os veículos que estão no alcance de cada veículo durante o tempo analisado.

4.4.3 Quantidade de dados de localização utilizados

O algoritmo desenvolvido, apesar de usar todo o banco de dados para identificar as regiões com mais possibilidade de transferência de dados, utiliza somente as 100 maiores regiões encontradas nas análises posteriores.

Ao selecionar somente as regiões mais movimentadas da cidade, os dados de localizações de veículos e determinados veículos são desprezados nas análises. Para mensurar essa característica, as Figuras 4.4 e 4.5 mostram a porcentagem de dados presentes nos 100 clusters em relação à totalidade dos bancos de dados e variando com o raio de análise escolhido.

A Figura 4.4 mostra a quantidade de dados de localização de veículos, variando de próximo a 10% das localizações – ao usar o raio de 50 metros – até 65% das localizações – com raio de 1.000 metros.

De forma semelhante, a Figura 4.5 mostra a porcentagem de veículos agrupados nos clusters selecionados em relação ao total de veículos presentes no banco de dados. Esse dado mostra que ao utilizar clusters de 1.000 metros praticamente todos os veículos fazem parte da rede. Ao utilizar clusters de 50 metros, pouco menos de

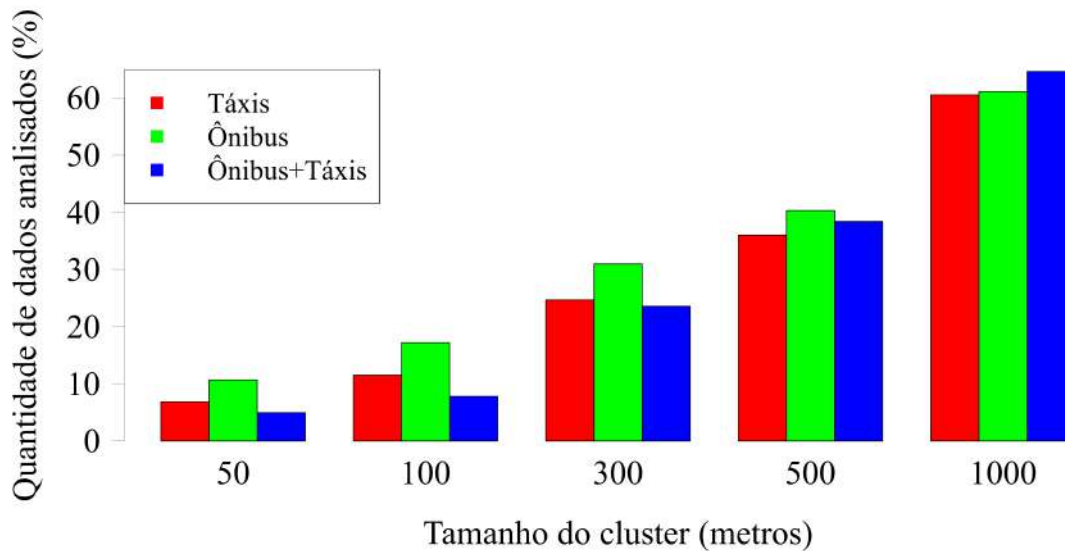


Figura 4.4: Quantidade de dados de localização analisados pelo algoritmo.

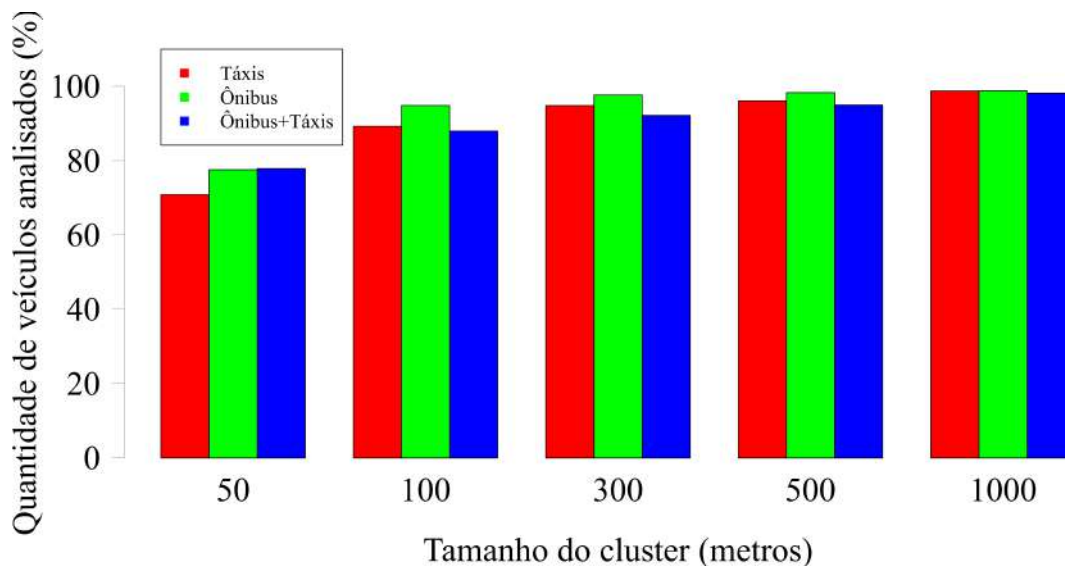


Figura 4.5: Quantidade de veículos analisados pelo algoritmo.

80% dos veículos são utilizados na rede modelada, ainda assim uma quantidade significativa dos veículos.

4.5 Conclusão

Este capítulo apresentou as metodologias utilizadas para a análise dos dados de deslocamento de veículos do sistema de transporte público apresentados no Capítulo 3. Neste capítulo foi estabelecido o cenário utilizado para a troca de dados do sistema e foram apresentadas as ferramentas usadas para análise de grandes massas de dados, os métodos de transmissão de dados aplicados ao cenário proposto e foi exibido o algoritmo de cálculo de viagens criado.

No Capítulo 5 é realizada uma análise estrutural dos dados dos ônibus e dos

táxis, observando a dinâmica da rede formada por veículos agrupados em clusters. A análise estrutural emprega o Algoritmo 1 para determinar o comportamento dos veículos, analisando-os em busca de padrões de comportamento da rede veicular.

Capítulo 5

Análise Estrutural

Este capítulo tem como objetivo realizar a análise dos dados obtidos ao se aplicar o Algoritmo 1 de clusterização descrito no Capítulo 4, Seção 4.4. As métricas obtidas com o algoritmo, a partir da localização dos ônibus e táxis, são utilizadas para realizar uma análise da estrutura de deslocamento dos veículos pela cidade do ponto de vista dos pontos de interesse selecionados, que são os 100 clusters mais movimentados da cidade. Este capítulo responde a primeira pergunta realizada na introdução: Como as regiões da cidade estão interligadas em relação ao deslocamento dos veículos?

Este capítulo está dividido em duas partes que tratam: da análise da dinâmica da rede na Seção 5.1, onde é realizada a primeira análise dos dados, buscando identificar a estrutura da rede estudada; e da análise da rede complexa formada pelas viagens entre pares de clusters na Seção 5.2.

No Capítulo 6 os dados obtidos serão analisados sob a óptica de uma rede veicular para transporte de dados, aplicando as restrições das tecnologias de transmissão já identificadas.

5.1 Dinâmica da rede

A dinâmica da rede corresponde às interações entre os diversos veículos e clusters da cidade. A partir dos dados processados pelo algoritmo são determinados 5 tipos de informações: localização dos clusters; número diário de veículos que trafegam por cada cluster; número de deslocamentos entre pares de clusters; tempo médio das viagens entre pares de clusters; e tempo de permanência dos veículos em cada cluster. Esses dados têm a capacidade de mostrar o comportamento dos veículos, identificando padrões de comportamento de deslocamento.

Os cálculos a seguir são realizados com os parâmetros de tamanho dos clusters de 300 metros de lado, analisando os 100 maiores clusters.

5.1.1 Número de atualizações por cluster

Inicialmente, todos os veículos do banco de dados são associados às células. Para a análise de cluster, somente as 100 maiores células (chamadas de clusters) são selecionadas para análise. A Figura 5.1 mostra o número de veículos que reportaram sua localização nos 100 maiores clusters ao longo de um intervalo de 24 horas, ou seja, o número de veículos de cada cluster. A Figura 5.1 mostra em verde (ônibus+táxis) a quantidade de atualizações por cluster quando são utilizados todos os veículos para a análise.

O cálculo é realizado a partir de 3 bancos de dados distintos: de ônibus, de táxis e um banco de dados as informações de ônibus e táxis combinadas. É possível observar que a quantidade de veículos em cada cluster é diferente nos 3 casos e que utilizar ambos os veículos para calcular os cluster tem um resultado diferente de se analisar e juntar o resultado dos ônibus e táxis de forma separada, já que os clusters são posicionados em locais distintos dependendo dos dados de entrada (ônibus, táxis ou ônibus e táxis). Logo, o cálculo para todos os veículos (verde) não é o simples somatório das atualizações dos ônibus (azul) com os táxis (vermelho).

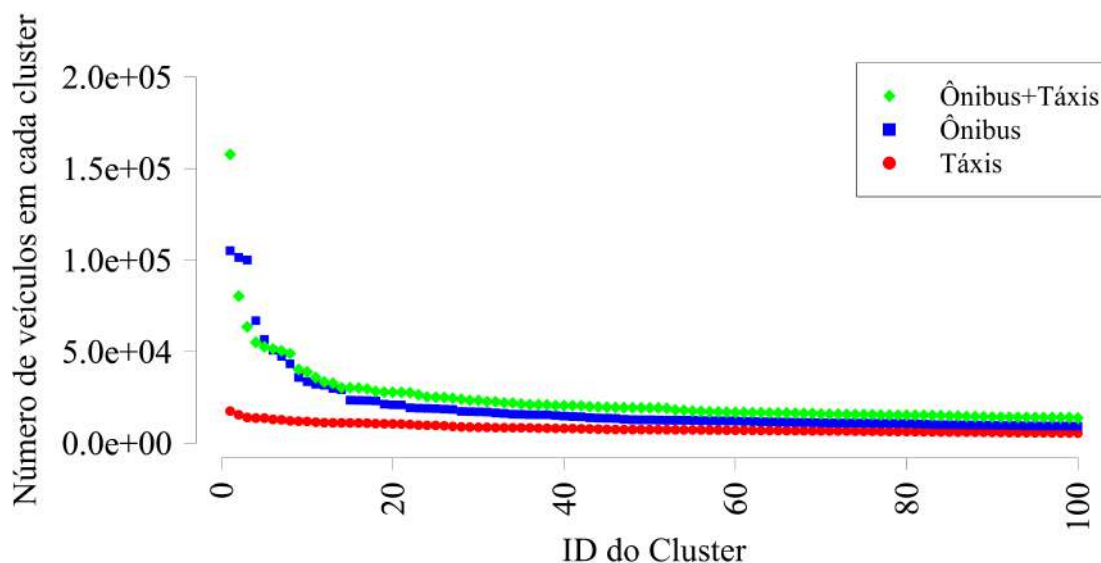


Figura 5.1: Quantidade de atualizações por cluster.

Todos os gráficos têm o mesmo comportamento: o primeiro cluster (1) de cada um tem uma elevada quantidade de veículos que informam a sua localização naquela região. Esta quantidade decai até que os últimos clusters analisados, que tem milhares de posições informadas. É possível observar que o número de veículos participantes modifica o número final de atualizações por cluster. Contudo, isso não é tão significativo nos clusters mais movimentados, pois entre os 6 primeiros clusters a quantidade de clusters formados exclusivamente por ônibus é maior do que a quantidade de clusters formados por ônibus e táxis, exceto no primeiro cluster.

5.1.2 Mapa dos clusters

O mapa da cidade do Rio de Janeiro é exibido nas próximas figuras com a localização das células de igual tamanho na cidade do Rio de Janeiro, em forma de grade, e com os 100 clusters em destaque. As Figuras 5.2, 5.3 e 5.4 mostram células de 300 metros de lado na cidade do Rio de Janeiro, para ônibus, táxis, ou ônibus e táxis, respectivamente. Um veículo está em uma determinada célula se o dado obtido do GPS estiver dentro dos limites de uma determinada célula.

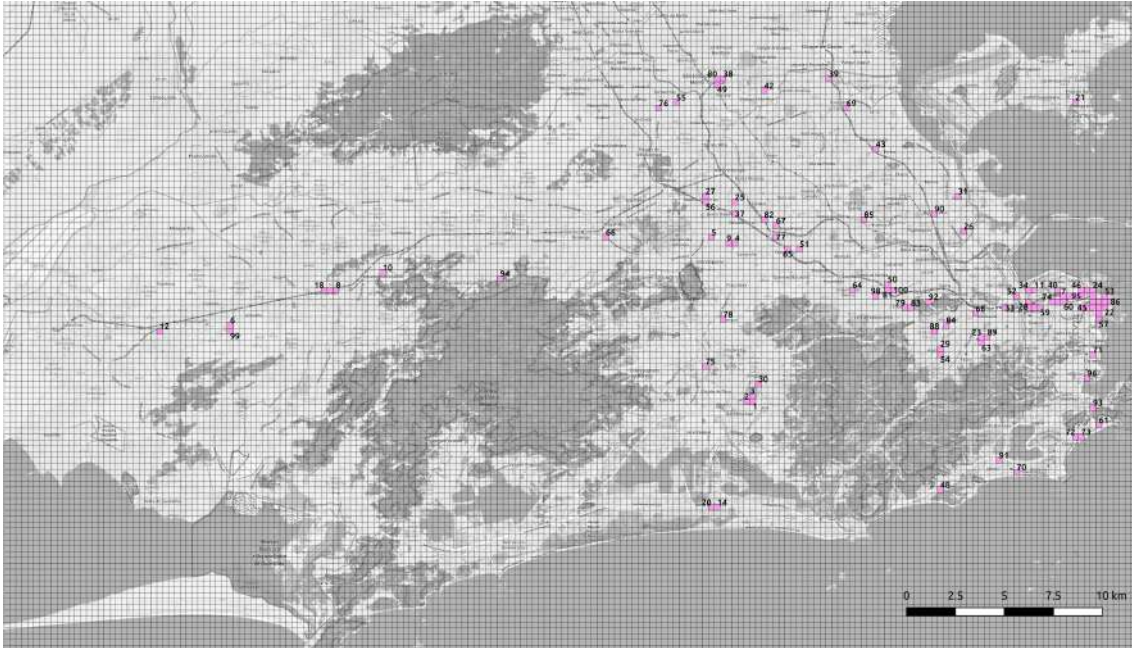


Figura 5.2: Mapa com células de 300 m e 100 maiores clusters formados por ônibus.

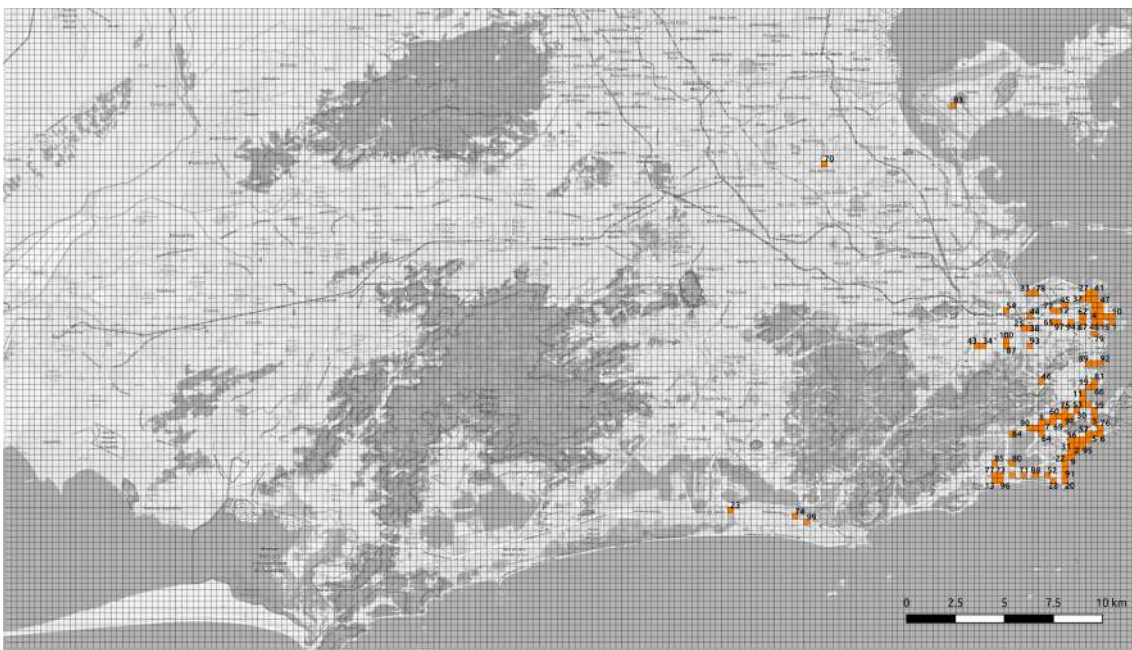


Figura 5.3: Mapa com células de 300 m e 100 maiores clusters formados por táxis.

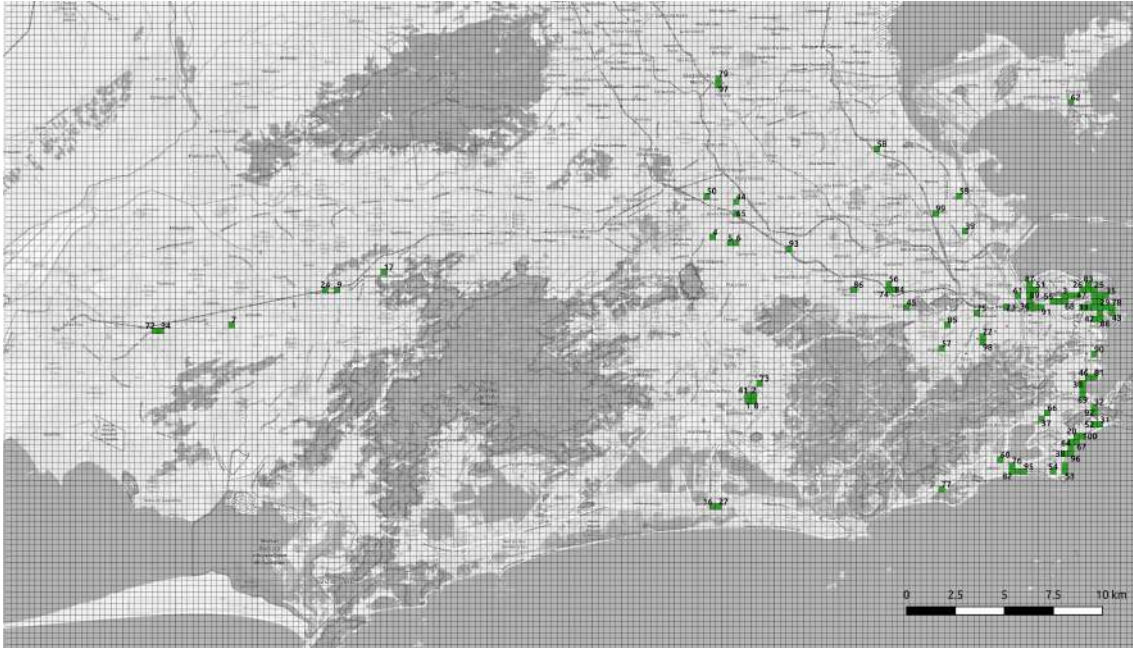


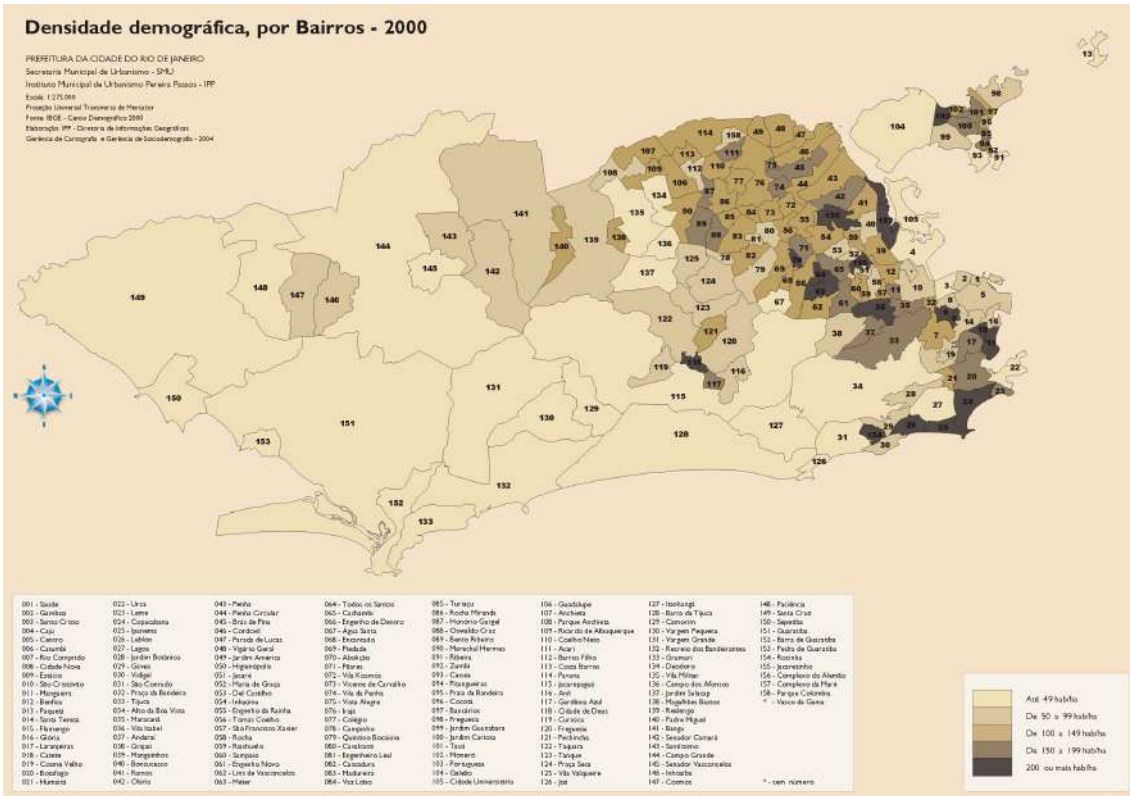
Figura 5.4: Mapa com células de 300 m e 100 maiores clusters formados por ônibus e táxis.

É possível observar entre os três mapas a diferença de localizações dos 100 maiores clusters. Quando o cluster é formado somente por dados de ônibus os clusters são bastante espalhados pela cidade, com uma concentração maior no centro da cidade (porção esquerda-central da cidade). Os clusters formados exclusivamente por táxis têm uma forte tendência de concentração somente no centro e na zona sul da cidade (porção esquerda-inferior da cidade). Ao juntar os dados de ambos os veículos e executar o algoritmo de clusterização, as tendências de ambos os veículos são mescladas, contendo clusters espalhados pela a cidade e um número maior de clusters na zona sul da cidade.

Aqui é importante destacar que a zona sul possui uma densidade demográfica maior do que as outras áreas da cidade, como mostra a Figura 5.1.2 [68], e é onde estão localizadas as famílias de mais alta renda, que têm a possibilidade de usar táxis de forma mais regular. Além disso, a área mais importante economicamente, onde a maioria das pessoas trabalha e onde o comércio se concentra, é a área central da cidade. Esses dois fatores podem ser utilizados para explicar a maior concentração de clusters localizados nessas áreas.

5.1.3 Distância entre pares de clusters

Formados os clusters, é importante descobrir a distância entre eles. Assim, mede-se a distância em linha reta entre clusters e forma-se uma matriz com as distâncias. Não foram escolhidas as distâncias entre clusters através das vias da cidade pois não é possível determinar o trajeto de todos os veículos que percorrem pares de clusters.



Fonte: Retirado de data.rio [68]

Figura 5.5: Mapa com a densidade demográfica da Cidade do Rio de Janeiro.

Além disso, os caminhos não são necessariamente iguais para todos os veículos que percorrem um par de clusters.

Para ilustrar, a Figura 5.6 exhibe a distância entre 100 clusters de todos os veículos (ônibus e táxis). A distância entre os clusters varia entre 300 m e cerca de 50 km, o máximo observado no Rio.

A Figura 5.7, por questão de melhor leitura, apresenta somente as 20 primeiras entradas das distâncias entre clusters. É possível observar a presença de clusters que poderiam ser otimizados ao escolher uma nova localização ou ao aumentar o raio de alcance do aparelho de rádio; no caso são os clusters que têm a distância de 0,3 km. Entretanto, esses clusters, que são fronteiros, não representam uma grande quantidade dos clusters formados, o que indica que o processo de clusterização é capaz de localizar clusters e viagens longas de forma satisfatória.

A Figura 5.8 apresenta o histograma das distâncias em quilômetros entre os pares de todos os 100 clusters. É possível observar que a distância máxima é menor ao utilizar os táxis do que os outros dois cenários. Quando se analisa os dados de localização de ônibus e táxis, o número de clusters com menor distância (até 10 km) é aumentado em relação aos ônibus. Já os clusters de 10 a 40 km têm o valor próximo, com ligeira vantagem quando são utilizados apenas os ônibus. A partir de

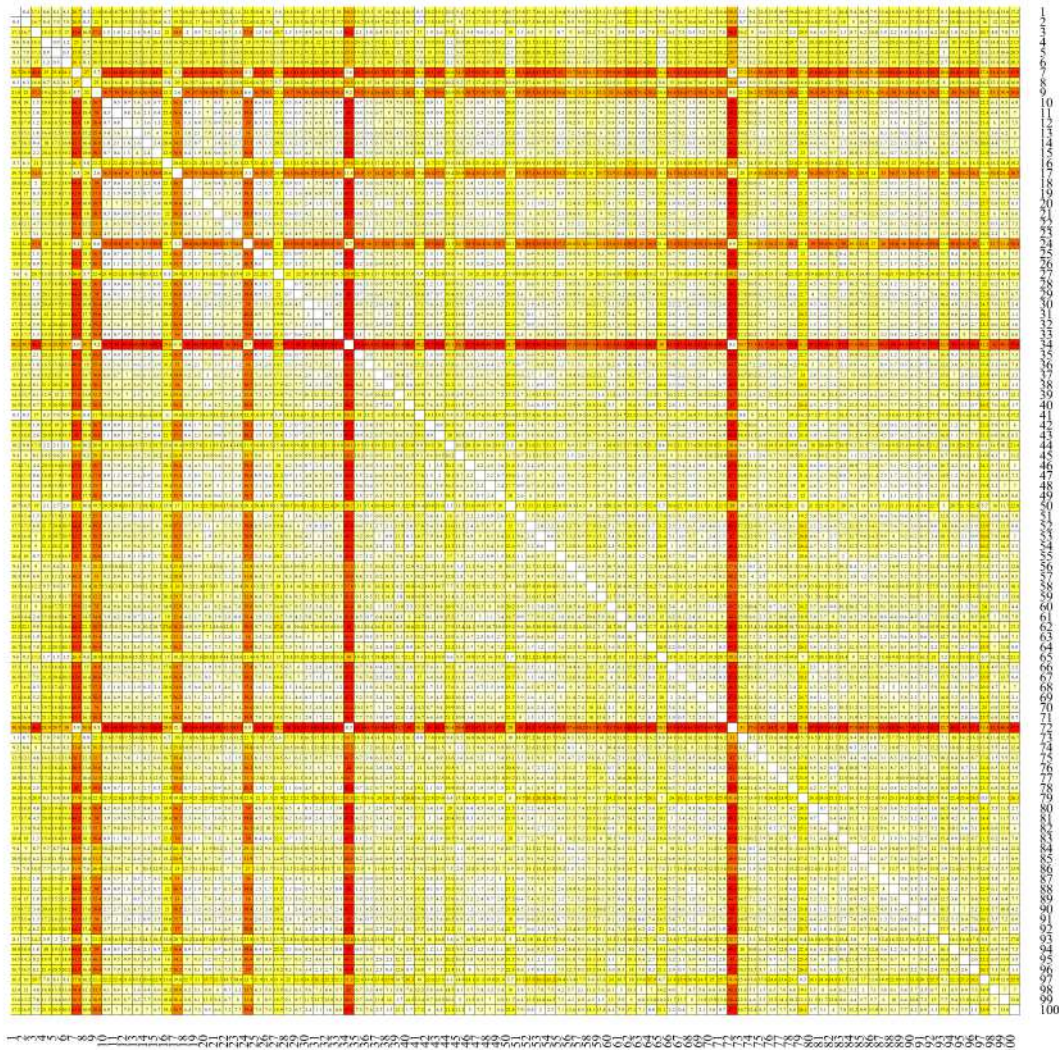


Figura 5.6: Distância (km) entre clusters de todos os veículos.

40 km não há diferença significativa no tamanho dos clusters.

A junção dos dados de táxis e ônibus faz diminuir a distância média entre pares de clusters. Porém, ao lembrar dos mapas das Figuras 5.2, 5.3 e 5.4, observa-se que existe uma maior quantidade de clusters na zona sul da cidade, próximos aos clusters do centro da cidade, o que faz com que seja diminuída a distância média.

	0.4	17.1	8.6	8.1	8.1	26.7	0.3	21.8	18.4	18.7	18.3	15.5	16.7	18.9	5.7	19.7	18.6	17.4	16.9	1
0.4		16.7	8.4	7.9	7.9	26.9	0.3	22	18	18.3	17.9	15.1	16.3	18.6	6.1	19.9	18.2	17	16.6	2
17.1	16.7		18.3	17.3	17	42.6	16.8	37.2	1.5	1.8	1.2	1.8	0.4	2.2	21	34.8	2	0.3	7.2	3
8.6	8.4	18.3		0.9	1.2	25	8.7	19.4	19.8	20.1	19.4	16.4	18	20.4	13.8	16.9	20.2	18.5	21.2	4
8.1	7.9	17.3	0.9		0.3	25.8	8.2	20.2	18.8	19.1	18.5	15.5	17.1	19.5	13.5	17.8	19.2	17.6	20.3	5
8.1	7.9	17	1.2	0.3		26.1	8.1	20.5	18.5	18.8	18.2	15.2	16.8	19.2	13.6	18.1	18.9	17.3	20	6
26.7	26.9	42.6	25	25.8	26.1		27	5.7	44.1	44.4	43.8	40.8	42.3	44.7	26.3	8.3	44.4	42.9	43.6	7
0.3	0.3	16.8	8.7	8.2	8.1	27		22	18.1	18.4	18	15.2	16.4	18.6	5.8	20	18.3	17.1	16.6	8
21.8	22	37.2	19.4	20.2	20.5	5.7	22		38.7	39	38.4	35.4	36.9	39.3	22.2	2.6	39	37.5	38.5	9
18.4	18	1.5	19.8	18.8	18.5	44.1	18.1	38.7		0.3	0.7	3.4	1.8	0.7	22.1	36.3	0.7	1.2	7	10
18.7	18.3	1.8	20.1	19.1	18.8	44.4	18.4	39	0.3		0.8	3.6	2.1	0.4	22.4	36.6	0.6	1.5	7	11
18.3	17.9	1.2	19.4	18.5	18.2	43.8	18	38.4	0.7	0.8		3	1.6	1.3	22.2	36	1.3	0.9	7.5	12
15.5	15.1	1.8	16.4	15.5	15.2	40.8	15.2	35.4	3.4	3.6	3		1.6	4	19.6	33	3.8	2.1	7.9	13
16.7	16.3	0.4	18	17.1	16.8	42.3	16.4	36.9	1.8	2.1	1.6	1.6		2.4	20.6	34.5	2.2	0.7	7	14
18.9	18.6	2.2	20.4	19.5	19.2	44.7	18.6	39.3	0.7	0.4	1.3	4	2.4		22.5	36.9	0.4	1.9	6.8	15
5.7	6.1	21	13.8	13.5	13.6	26.3	5.8	22.2	22.1	22.4	22.2	19.6	20.6	22.5		20.6	22.1	21.2	18.9	16
19.7	19.9	34.8	16.9	17.8	18.1	8.3	20	2.6	36.3	36.6	36	33	34.5	36.9	20.6		36.7	35.1	36.4	17
18.6	18.2	2	20.2	19.2	18.9	44.4	18.3	39	0.7	0.6	1.3	3.8	2.2	0.4	22.1	36.7		1.7	6.4	18
17.4	17	0.3	18.5	17.6	17.3	42.9	17.1	37.5	1.2	1.5	0.9	2.1	0.7	1.9	21.2	35.1	1.7		7.2	19
16.9	16.6	7.2	21.2	20.3	20	43.6	16.6	38.5	7	7	7.5	7.9	7	6.8	18.9	36.4	6.4	7.2		20
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	

Figura 5.7: Distância (km) entre os 20 maiores clusters de todos os veículos.

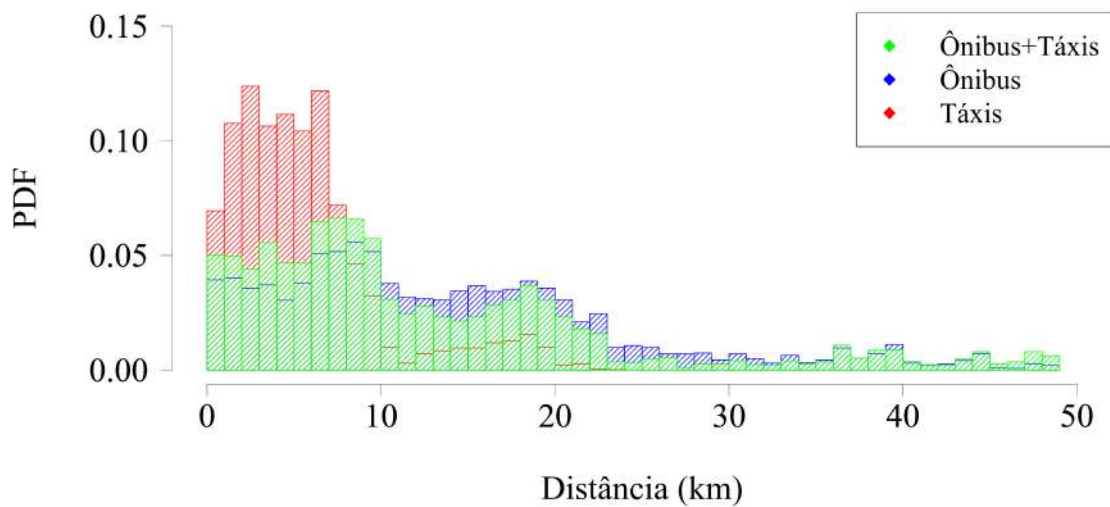


Figura 5.8: Histograma da distância (km) entre os clusters.

5.1.4 Viagens entre pares de clusters

A Figura 5.9 mostra a função de densidade de probabilidade (PDF) do número de viagens entre clusters. Em todos os casos, uma parte considerável dos pares de clusters exibe uma quantidade de viagens menor do que 500 viagens por dia, significando que ocorre uma viagem a cada 3 minutos ou mais. Entretanto, para todos os casos, a distribuição tem uma cauda longa, indicando que determinados pares de clusters apresentam uma quantidade elevada de viagens no período.

O número de clusters com pouquíssimas viagens é maior quando são utilizados somente ônibus. O número máximo de viagens utilizando somente táxis é o menor, tendo a cauda mais curta entre as distribuições, porém as viagens são distribuídas de forma mais uniforme. O número de viagens decai mais rápido quando se considera a rede com os ônibus e táxis ao ser comparado com as distribuições de somente ônibus ou somente táxis. Ao utilizar ambos veículos o número de pares de clusters com valores elevados é o maior.

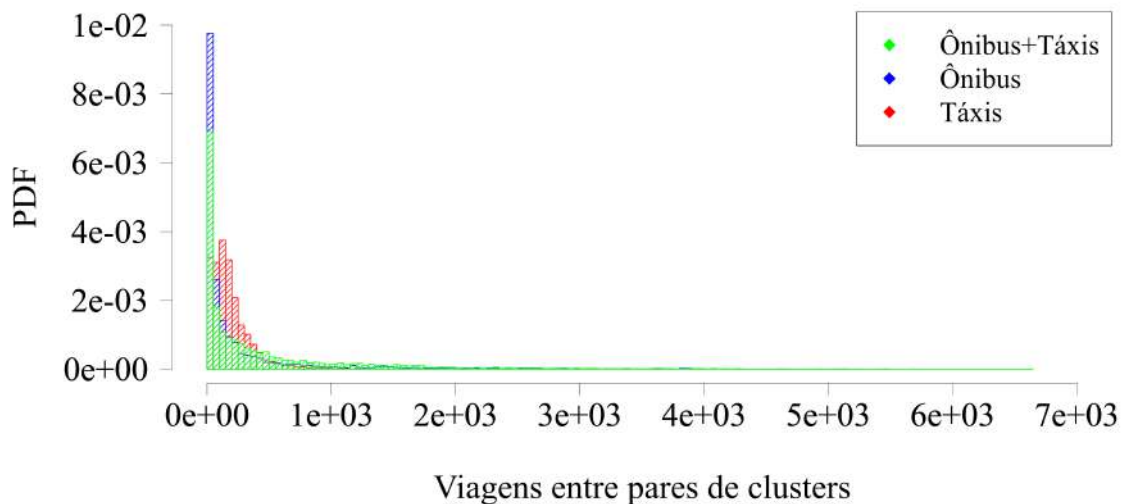


Figura 5.9: PDF do número de viagens entre pares de clusters.

A Figura 5.10 exibe a quantidade total de viagens de cada um dos 100 clusters. Nessa figura estão representadas as viagens quando são utilizados ônibus (círculo azul), táxis (quadrados vermelhos) e ônibus e táxis de forma conjunta (losango verde). A partir da análise da figura não é possível determinar um padrão para o número total de viagens. Clusters com grande quantidade de veículos que reportaram a sua localização naquele cluster, ou seja, clusters com número de identificação (ID) baixo apresentam poucas viagens. Clusters com um menor número de reportes de veículos, ID alto, apresentam muitas viagens.

O número de viagens de cada um dos clusters de táxis tem a variação menor do que os clusters de ônibus, além do valor total de viagens ser menor, o que pode ser

explicado pelas rotas fixas dos ônibus e das rotas não constantes dos táxis.

O caso específico dos pontos da Figura 5.10 na parte inferior esquerda, com ID alto e baixa quantidade de viagens, é ocasionado por esses pontos estarem localizados em garagens de empresas de ônibus, permanecendo um maior tempo nesse local, mostrando que esses locais podem ser utilizados em benefício das transferências de dados que podem ter latências elevadas.

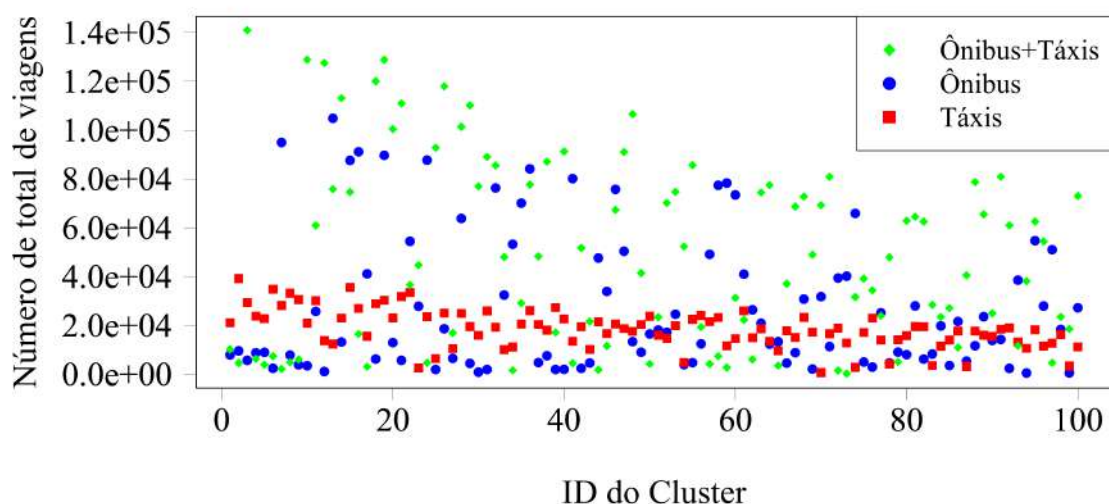


Figura 5.10: Número de viagens por ID do cluster.

A Figura 5.11 representa o número de viagens em relação com a distância dos ônibus (losango azul), táxis (losango vermelhos) e ônibus e táxis de forma conjunta (losango verde), ou seja, é o número de viagens de cada distância. No eixo X é exibida a distância entre pares de clusters, ou seja, a distância das viagens, e no eixo Y é exibido o número de viagens que tiveram determinada distância. Por exemplo, é identificado que existem 1.000 viagens entre um par de clusters, onde a distância é 1 km. Além desse, existe outro par de clusters com 5 viagens e a mesma distância de 1 km. Ambos os dados de viagens (5 e 1.000) são plotados sobre o valor de 1 km.

Ao correlacionar o número de viagens com a distância entre os clusters, observa-se que os aglomerados próximos geralmente apresentam mais viagens entre eles durante o dia, tanto para ônibus quanto para táxis, como mostra a maior intensidade de pontos com distâncias menores do que 10 km. A partir de 20 km existe um predomínio nas viagens de ônibus, porém de forma menos intensa que nas distâncias curtas.

5.1.5 Tempo médio de deslocamento entre pares de clusters

O tempo médio de deslocamento entre pares de clusters é determinado pela soma de todos os tempos de viagem entre cada par de clusters, divididos pelo número de

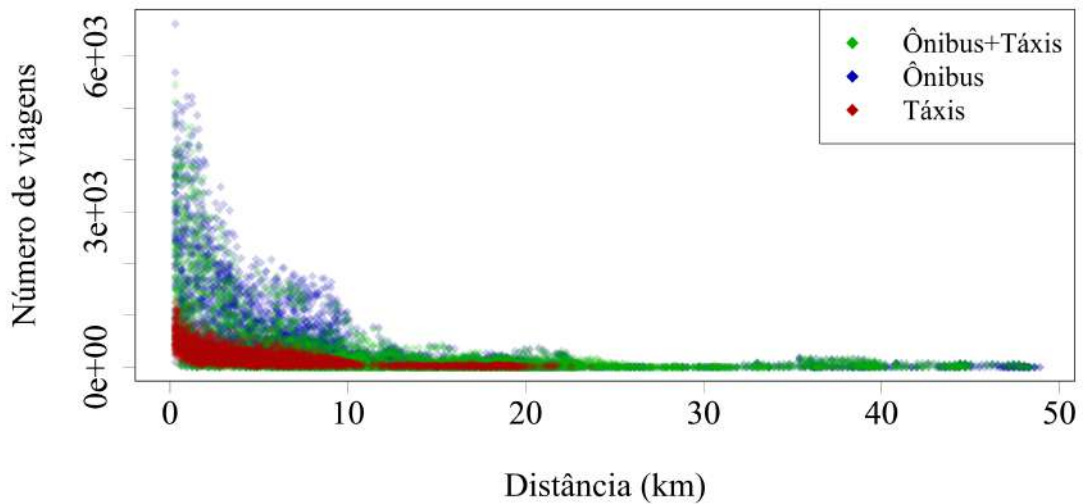


Figura 5.11: Número de viagens em relação à distância (km).

viagens que ocorrem entre esses clusters. A Figura 5.12 mostra a função de densidade de probabilidade (PDF) dos tempos de viagem entre os clusters, onde cada barra corresponde a um intervalo de 20 minutos.

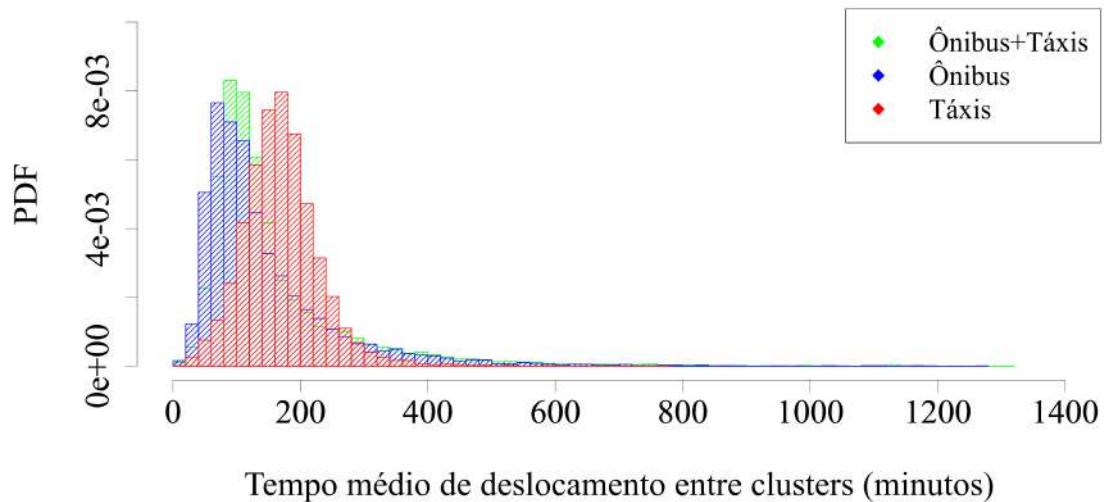


Figura 5.12: PDF do tempo médio de deslocamento entre pares de clusters.

A PDF do tempo de viagem entre os clusters dos ônibus apresenta um pico entre 60 e 80 minutos, enquanto o tempo de viagem entre clusters da PDF de táxis tem um pico entre 160 e 180 minutos. Uma das razões para essa diferença de tempo de pico de viagens entre clusters é explicada pelo fato dos ônibus possuírem uma trajetória predefinida, normalmente circular, já os táxis não apresentam essa característica. O número de viagens entre clusters de táxi, em média, é mais espalhado pelo tempo em comparação a ônibus, além de apresentar um número máximo de viagens menor, mostrando que as viagens entre pares de clusters detectadas não têm um trajeto tão

predeterminado quanto o dos ônibus. Os tempos médios das viagens entre pares de cluster ao usar ambos os veículos são predominantemente baixos, cerca de 100 minutos.

A Figura 5.13 apresenta a matriz de tempo de viagem médio, em minutos, entre os 20 primeiros clusters utilizando ambos os veículos. Ao observar a matriz de tempo de viagens é possível identificar que o tempo de viagem entre os pares de clusters é dependente da direção. Por exemplo, o par de clusters 17-20 requer um tempo de viagem de 236 minutos do cluster 17 ao 20 e 773 minutos do cluster 20 ao 17 usando táxis e ônibus.

NaN	47	272	877	154	136	NaN	19	NaN	231	170	215	277	227	249	89	NaN	233	254	254	1
155	NaN	251	NaN	207	186	NaN	47	NaN	166	156	266	209	268	290	76	NaN	259	289	217	2
206	208	NaN	112	151	110	156	218	121	70	102	68	86	58	98	127	135	64	68	69	3
573	NaN	160	NaN	54	73	NaN	NaN	NaN	182	283	191	140	167	184	72	NaN	176	181	147	4
396	646	176	154	NaN	53	NaN	756	NaN	163	286	153	134	150	136	187	NaN	153	141	498	5
267	474	140	72	84	NaN	NaN	555	NaN	138	269	122	121	137	150	197	28	138	113	344	6
NaN	NaN	415	NaN	NaN	NaN	NaN	NaN	420	542	229	363	221	296	345	196	NaN	521	253	241	7
37	31	282	NaN	500	474	NaN	NaN	NaN	232	170	217	386	236	288	142	NaN	234	245	289	8
NaN	NaN	135	NaN	NaN	NaN	138	NaN	NaN	255	139	175	112	127	147	90	36	174	113	186	9
198	184	50	112	124	130	346	208	157	NaN	53	66	96	76	69	133	155	42	72	65	10
148	207	51	189	188	141	173	178	129	59	NaN	57	110	80	107	157	148	88	86	111	11
178	210	46	100	121	110	175	176	171	46	130	NaN	90	65	108	129	245	50	47	67	12
249	284	73	139	224	145	281	252	142	79	83	83	NaN	61	90	127	148	89	69	83	13
198	197	62	112	157	119	142	218	122	63	95	63	106	NaN	93	287	130	72	63	105	14
217	204	47	104	167	128	254	251	158	74	40	46	107	72	NaN	188	258	64	68	122	15
84	111	105	150	264	261	244	125	200	143	106	125	110	262	114	NaN	316	142	116	126	16
NaN	NaN	218	NaN	NaN	19	NaN	NaN	34	415	198	461	168	192	434	406	NaN	539	246	236	17
194	241	61	105	125	118	310	197	192	70	64	64	91	80	61	141	219	NaN	61	62	18
213	219	41	111	152	101	156	203	143	61	82	69	109	60	90	152	224	67	NaN	80	19
187	198	72	247	231	288	NaN	233	133	92	159	85	112	98	111	126	773	103	74	NaN	20
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	

Figura 5.13: Matriz do tempo médio de deslocamento entre os 20 primeiros clusters de ambos os veículos.

As Figuras 5.14 e 5.15 representam o tempo médio das viagens entre clusters em relação à distância. No eixo X é exibida a distância entre pares de clusters, ou seja, a distância das viagens, e no eixo Y é exibido o tempo médio de viagens que tiveram determinada distância.

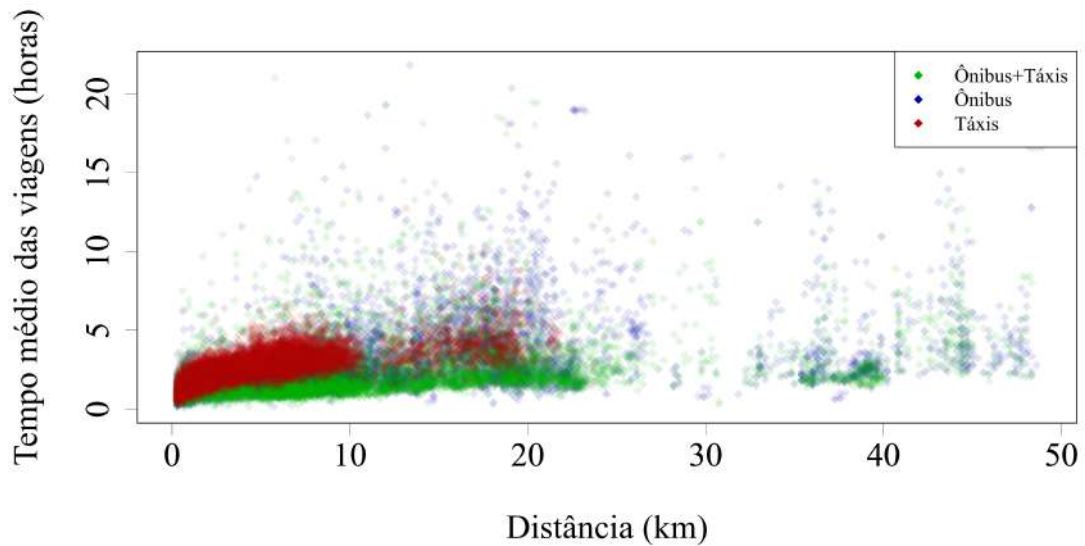


Figura 5.14: Tempo médio das viagens entre clusters em relação à distância (km).

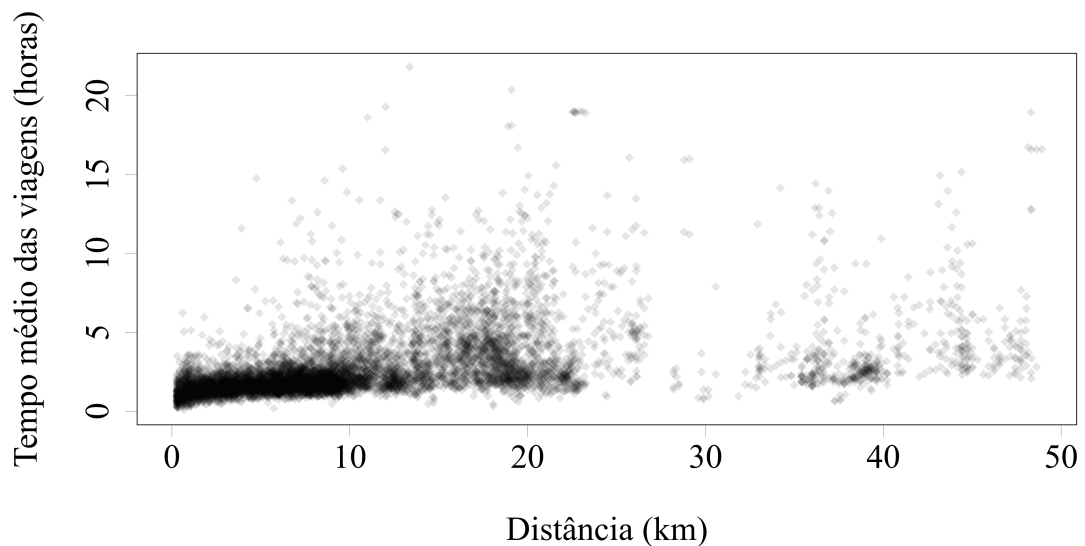


Figura 5.15: Tempo médio das viagens de ambos os veículos entre clusters em relação à distância (km).

A Figura 5.14 mostra, em vermelho, as viagens dos táxis, cujo tempo de viagem aumenta progressivamente de acordo com a distância percorrida. Não são identificadas viagens excessivamente longas, ou maior do que 10 horas, e viagens longas, maiores do que 20 km. A Figura 5.15 mostra o tempo médio das viagens com ônibus e táxis simultaneamente.

As figuras mostram que o tempo mínimo de viagem, ou distância do eixo X ao gráfico, aumenta de acordo com a distância, o que mostra que distâncias maiores necessitam de viagens mais longas. Isso implica que clusters mais distantes tem uma maior latência para a troca de dados, como esperado.

De forma geral é observado que pares de clusters próximos tem uma maior quantidade de viagens, sendo elas com baixo tempo médio de viagens. Essa característica

mostra que é possível transferir uma grande quantidade grande de informações entre clusters próximos. Além disso existem viagens com tempo baixo entre clusters distantes, mostrando que mesmo clusters distantes podem ter um tempo baixo de disseminação de dados. Por último, apesar de existirem viagens com tempo elevado entre clusters, existe uma concentração maior nas viagens de menor tempo do que nas viagens de maior tempo, mostrando que a rede formada pode ter uma baixa latência na transmissão de informações entre clusters.

O menor número de viagens entre 25 e 35 km é explicado pela geografia da cidade do Rio de Janeiro, que é tem seu território com forma alongada e achatada. Os clusters tem uma grande tendência de formação na extremidade direita da cidade, no centro e na zona sul da cidade. Os clusters formados na outra extremidade da cidade, zona norte e oeste da cidade, são menos numerosos, levando a poucos pares de clusters com distância elevada. Por essa formação, existem poucos clusters com distâncias entre 25 e 35 km.

5.1.6 Tempo médio de permanência nos clusters

A Figura 5.16 mostra a função de densidade de probabilidade (PDF) do tempo médio que os veículos reportam que permanecem na região de cada um dos clusters de origem. É importante mencionar que o tempo de permanência é medido individualmente para as viagens diferentes (de A para B e de B para A).

A distribuição do tempo de permanência para ônibus tem cauda longa, ou seja, alguns clusters têm um tempo de permanência dos veículos muito longo (75.000 minutos ou 1.250 horas) no período de 24 horas, lembrando que o tempo de permanência é o somatório do tempo de permanência de todos os veículos que estiveram naquele cluster durante as 24 horas. Por outro lado, há uma elevada concentração em valores próximos a zero minutos, o que pode ser explicado pelo fato de existirem clusters com uma menor quantidade de veículos, clusters que se encontram em regiões mais afastadas, tendo uma menor diversidade de viagens para outros clusters. Além disso, um ônibus pode simplesmente cruzar um cluster, não permanecendo por um período de tempo. Poucos ônibus permanecem muito tempo dentro de um cluster específico. Estes clusters geralmente correspondem a lugares como garagens ou pontos de final de linha.

A cauda da distribuição dos táxis é a mais curta e o pico da distribuição ocorre em tempos de permanência maiores do que o dos ônibus. Em outras palavras, uma menor quantidade de clusters têm tempos de permanência muito curtos. Ao considerar-se a rede que combina ônibus e táxis, há uma redução do número de clusters com tempo de permanência baixo e a distribuição mantém uma cauda longa, apesar de menor do que a dos ônibus.



Figura 5.16: PDF do tempo de permanência em cada clusters.

O tempo de permanência está relacionado com a capacidade da infraestrutura necessária para transportar os dados entre os veículos, os pontos de acesso e a Internet. Tempos de permanência mais longos significam que uma maior quantidade de veículos se comunicam com o ponto de acesso do cluster, aumentando potencialmente a quantidade de dados transferidos. O equipamento de rádio e a conexão à Internet de cada cluster devem ser capazes de lidar com a quantidade prevista de veículos e de dados, assim, o tempo de permanência impacta nos requisitos da rede sem-fio e da conexão à Internet, ou backbone, devendo ser utilizado para calcular a implementação da infraestrutura.

5.2 Análise da rede complexa

Ao analisar o número de viagens entre clusters é possível identificar padrões, como a importância de uma região da cidade. A ideia é revelar as regiões com um alto volume de veículos e, conseqüentemente, maiores oportunidades de transporte de dados. Para analisar a rede complexa o primeiro passo necessário é criar um grafo onde cada vértice é uma região geográfica da cidade e as arestas representam as viagens entre as regiões.

5.2.1 Viagens entre pares de clusters

Para uma melhor análise, a Figura 5.17 mostra um grafo direcionado ponderado dos 100 maiores clusters identificados a partir do número de viagens entre clusters utilizando ambos os veículos. Os vértices são os clusters, numerados de 1 (contendo o maior número de veículos) a 100 (contendo o menor número de veículos). O

tamanho do vértice (diâmetro do círculo) representa o valor da intermediação, ou *betweenness centrality*, de cada um deles. As 8.904 arestas representam viagens de veículos entre clusters e a cor da aresta e a espessura representam o número de viagens, variando do azul (maior número de viagens) ao cinza (menor número de viagens).

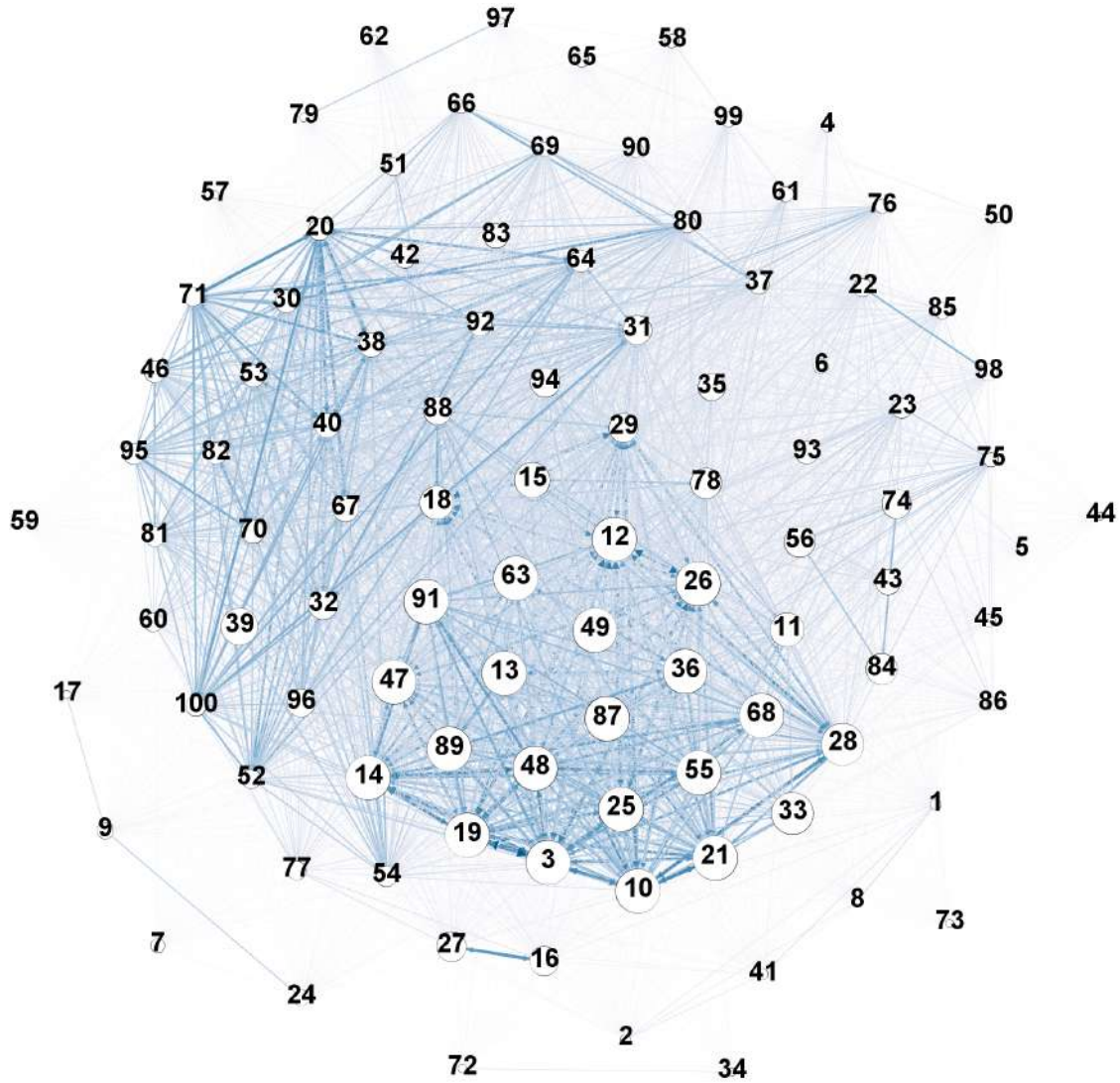


Figura 5.17: Grafo representativo do número de viagens entre clusters de 300 metros.

A primeira observação interessante é que o grafo possui um único componente conectado: todos os vértices são acessíveis, diretamente ou através de múltiplos saltos. Como um grafo direcionado ponderado, ele tem dois graus: o grau de entrada e grau de saída. O menor grau de entrada é 29, enquanto o maior é 99. O menor grau de saída é 28 e o maior é 99. Dezenove por cento dos vértices têm o grau máximo (198) e apenas 2% dos vértices têm um grau inferior a 100. O grau médio do grafo é 178,08, significando que a rede é altamente conectada. Por outro lado, o diâmetro do grafo é 2, o que significa que todos os vértices podem ser alcançados em pelo

menos 2 saltos. O comprimento médio dos caminhos na rede é de 1,1 saltos.

Outra métrica importante é o número de caminhos mais curtos que atravessam um vértice, ou intermediações (*betweenness centrality*). O tamanho dos vértices na Figura 5.17 é proporcional à intermediação do vértice. Os vértices mais centrais são importantes porque um grande número de veículos os atravessam. Assim, esses vértices são bons candidatos para compor a espinha dorsal da rede. Os vértices 3, 10, 12, 13, 14, 19, 21, 25, 26, 36, 47, 48, 49, 55, 63, 68, 87, 89 e 91 têm o valor de intermediação de 21,76. É importante destacar que os vértices com o maior número de veículos não são necessariamente os mais centrais em relação às viagens entre clusters.

Como qualquer cluster pode ser alcançado em até 2 saltos, para transferir uma informação entre 2 locais da cidade (clusters), sem a ajuda de qualquer outra rede, precisamos de no máximo 1 cluster para retransmitir a informação para outro veículo. O algoritmo de roteamento necessário é bastante simples, já que apenas 2% de clusters têm conexão muito limitada com outros clusters e o caminho médio entre clusters é 1,1, necessitando de menos retransmissões na rede. Os clusters com alta centralidade de interação são ótimos candidatos para atuar como pontos de retransmissão neste cenário, já que a maioria dos caminhos entre os vértices passa por eles.

Na rede modelada, os veículos são responsáveis fisicamente por transportar dados, ou seja, os veículos são os nós móveis que transportam dados de um ponto para outro. Ao considerar o tempo de permanência de um veículo em um cluster como o tempo que ele passa no alcance de um equipamento de rádio localizado no cluster, o tempo de permanência se traduz em uma quantidade de dados transferidos. O tempo médio de viagem é usado para modelar o fluxo de dados na rede.

5.3 Conclusão

Este capítulo apresentou a análise dos dados obtidos ao se aplicar o Algoritmo 1 de clusterização descrito na Seção 4.4, realizando uma análise da estrutura de deslocamento dos veículos pela cidade do ponto de vista dos pontos de interesse selecionados, que são os 100 clusters mais movimentados da cidade.

Este capítulo respondeu a primeira pergunta realizada na introdução: Como as regiões da cidade estão interligadas em relação ao deslocamento dos veículos? Foi identificado que táxis e ônibus têm padrão de movimentação distinto, o primeiro com rotas do tipo origem-destino e o segundo seguindo rotas predefinidas por grandes vias. Além disso, os táxis têm a sua maior concentração nas áreas de maior densidade demográfica da cidade. Ônibus, na forma analisada, realizam viagens mais longas do que táxis, chegando a ter o dobro da distância máxima. A análise também mostra que as garagens dos ônibus são um local com alto tempo de permanência, podendo

ser utilizadas em benefício das transferências de dados capazes de suportar elevadas latências.

No Capítulo 6 os dados obtidos serão analisados sob a óptica de uma rede veicular para transporte de dados, aplicando as restrições das tecnologias de transmissão já identificadas para determinar a latência da rede e nos cálculos da capacidade da rede.

Capítulo 6

Análise de Capacidade

A partir dos dados coletados no Capítulo 5 é analisada a quantidade de dados que podem ser transportados na rede. Para isso assume-se que cada veículo é capaz de armazenar e transmitir dados através de equipamentos de rádio a bordo do veículo. Quando um veículo se aproxima de um cluster e entra em alcance do equipamento de rádio instalado no cluster, uma transmissão de dados por meio sem-fio é realizada. Este capítulo responde a segunda pergunta realizada na Introdução: Quanta informação é possível transmitir entre essas regiões em função do tempo?

Os dados a serem transmitidos podem ter várias aplicações, não levadas em consideração aqui. Por exemplo: os veículos podem coletar dados de outros dispositivos que se encontram nos seus trajetos, carregando-os para um local central, através da Internet, quando em contato com um cluster. Além disso, veículos podem gerar dados através de sensores, como condições ambientais ou das vias. Por último, os veículos podem ser utilizados somente para transporte de dados entre clusters, seja somente para enviar um dado para outro veículo em outro cluster ou para transporte de grandes dados entre localizações da cidade. Dessa forma, nesta tese, focamos na capacidade de dados que podem ser transportados pela rede veicular e não como esses dados são gerados.

Neste capítulo é analisada a capacidade de transmissão de dados dependente da tecnologia na Seção 6.1, sendo um aspecto chave para avaliar as aplicações que a rede veicular pode suportar. Outra medida importante é a latência da transmissão, explorada na Seção 6.2.

6.1 Capacidade de transferência de dados

Para inferir a capacidade de transferência de dados são avaliadas diferentes tecnologias de redes sem-fio. Nesta tese, como apresentado no Capítulo 4, Seção 4.3, varia-se o raio de contato e utiliza-se um fator de multiplicação α para simular determinada tecnologia de transmissão sem-fio.

A Tabela 4.7, apresentada anteriormente, fornece o alcance e a capacidade de transmissão típicos de cada tecnologia sem-fio. Baseadas nessa tabela, as análises seguintes apresentam o tamanho de uma célula com as dimensões de: 50 metros, 100 metros, 300 metros, 500 metros e 1.000 metros.

A Figura 6.1 apresenta a quantidade de veículos que reportaram sua localização nos 100 maiores clusters ao longo de um intervalo de 24 horas, ou seja, o número de veículos de cada cluster. São apresentados os valores para 5 tamanhos de cluster, onde o asterisco preto apresenta dados de clusters de 1.000 metros, o triângulo rosa apresenta dados de cluster de 500 metros, o losango verde apresenta dados de clusters de 300 metros, o quadrado vermelho apresenta dados de clusters de 100 metros e o círculo azul apresenta dados de clusters de 50 metros. Ao longo deste capítulo, o padrão de símbolos e cores citado é utilizado nas figuras para representar as distâncias de clusters destacadas e, ao menos que seja explicitamente indicado, em todas as análises é utilizado o banco de dados composto por ambos os veículos, ônibus e táxis, de forma simultânea.

Como esperado, o aumento do alcance da tecnologia de transmissão sem-fio causa um aumento no número de contatos de veículos com os clusters. O maior dos clusters recebe cerca de 250 mil relatórios de localização de veículos em um dia, para o cenário de alcance de rádio de 1.000 metros, enquanto recebe 160 mil em cenários de 300 metros e 50 mil em 50 metros de alcance, respectivamente.

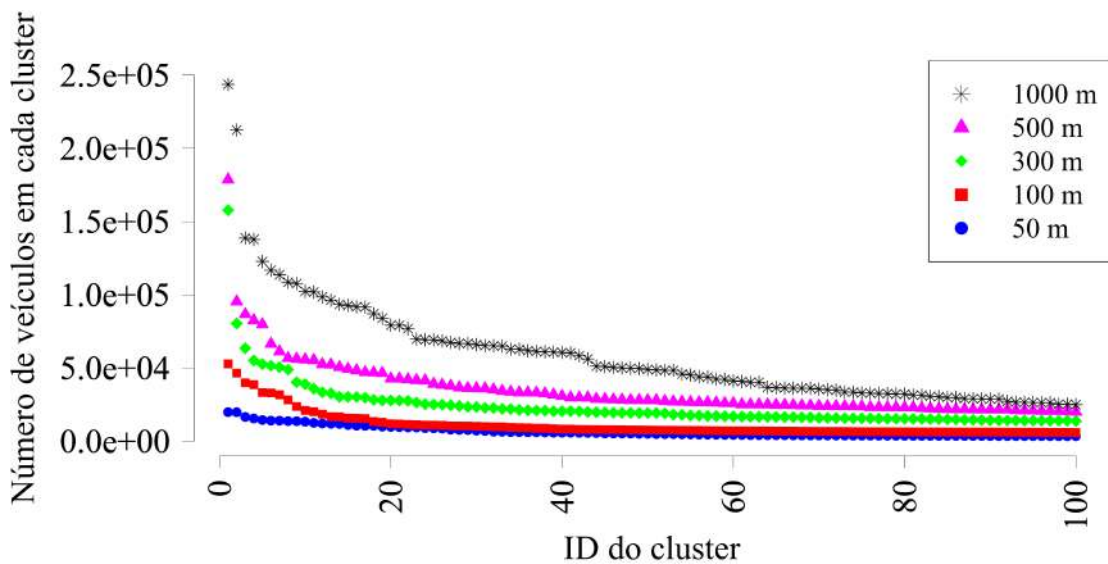


Figura 6.1: Quantidade de veículos que reportam a sua localização em cada cluster.

A Figura 6.2 mostra o somatório de tempo em que todos os veículos permanecem no alcance dos clusters, durante o dia analisado, considerando todos os 100 clusters. Este somatório corresponde ao tempo total que os veículos podem usar para carregar e descarregar dados para a rede em um dia. A primeira coluna de cada grupo mostra apenas o tempo total dos táxis, a coluna do meio, apenas os ônibus, enquanto a

última coluna mostra o tempo total considerando ambos os veículos. Cada grupo de barras corresponde a uma rede sem-fio diferente, de alcance de transmissão de 50 a 1.000 metros.

O tempo de permanência aumenta com o alcance do rádio sem-fio, o que é esperado, uma vez que os veículos tendem a permanecer mais tempo no alcance do equipamento de rádio dos clusters.

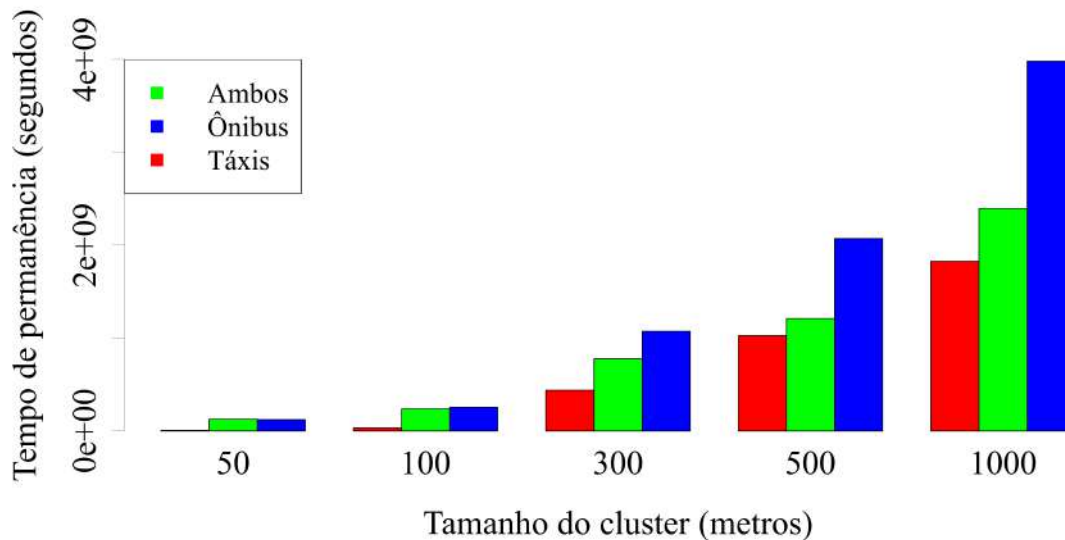


Figura 6.2: Tempo total de permanência dos veículos nos clusters, por tipo de veículo.

Para calcular a quantidade total de dados que podem ser transferidos em um dia, é necessário apenas usar o fator de multiplicação α correspondente a cada tecnologia, de acordo com a Tabela 4.7. Como o fator α é dado em megabits por segundo ou Mbps, basta multiplicar pelo tempo de permanência para termos a quantidade de dados transportados em megabits.

A Tabela 6.1 mostra, em Terabytes, a quantidade total de dados que podem potencialmente ser enviados através considerando uma conexão de 1 Mbps de cada veículo com o ponto de acesso no cluster. Os dados estão separados por tipo de veículos, ônibus, táxis e ônibus e táxis de forma simultânea e por tamanho do cluster, 50, 100, 300, 500 e 1.000 metros. Essa baixa capacidade de transmissão de dados foi escolhida por ser um valor realizável pelas tecnologias de transmissão sem-fio que atualmente podem ser utilizadas nos veículos.

Por exemplo, a rede formada pelos dois tipos de veículos (ônibus e táxis) e pontos de acesso nos clusters – utilizando tecnologias de transmissão entre os veículos e os pontos de acesso com alcance de 1.000 m e banda de 1 Mbps – é capaz de transferir até 4×10^9 megabits, ou 5×10^8 megabytes (500 TB), em um período de 24 horas. No outro extremo, clusters de 50 metros produzem um tempo de permanência total de 1.2×10^8 segundos. Assumindo também conexões com largura de banda de 1 Mbps,

Tabela 6.1: Potencial de transferência de dados por dia (Terabytes).

Alcance (metros)	Dados transferidos (TB)		
	Táxis	Ônibus	Ambos
50	0,75	15,75	15,19
100	4,29	29,71	32,01
300	54,88	97,15	134,07
500	128,22	151,38	258,92
1.000	228,40	299,38	497,61

os dados totais descarregados por todos os veículos atingem 15 TB. Em um outro exemplo, menos restritivo, utilizando conexões de 10 Mbps e alcance de 300 m é possível transferir 1,34 PB de dados por dia.

Observando a Figura 6.2, o uso de táxis e ônibus em conjunto melhora a capacidade da rede quando o alcance da rede sem-fio considerado é de 300 metros ou mais, principalmente devido ao maior número de veículos. Entretanto, a capacidade da rede ao usar os dois veículos simultaneamente não é simplesmente a soma das redes de táxis e ônibus isoladamente. Esse comportamento ocorre devido aos diferentes padrões de movimentação dos veículos, que leva à formação de clusters em locais distintos. As regiões com mais movimento de táxis ou ônibus diferem e, à medida que o número de clusters é fixado em 100, os locais do cluster ao usar ambos os veículos podem não estar nas regiões ideais para ônibus ou táxis sozinhos. Curiosamente, nas análises com o alcance menor do que 300 m usar ambos os veículos é pior do que usar apenas os ônibus. A rede formada somente por táxis também, em todos os casos, é a que tem a menor capacidade de transmissão, entretanto, com alcances de 300 m ou mais, a rede tem a capacidade semelhante aos ônibus.

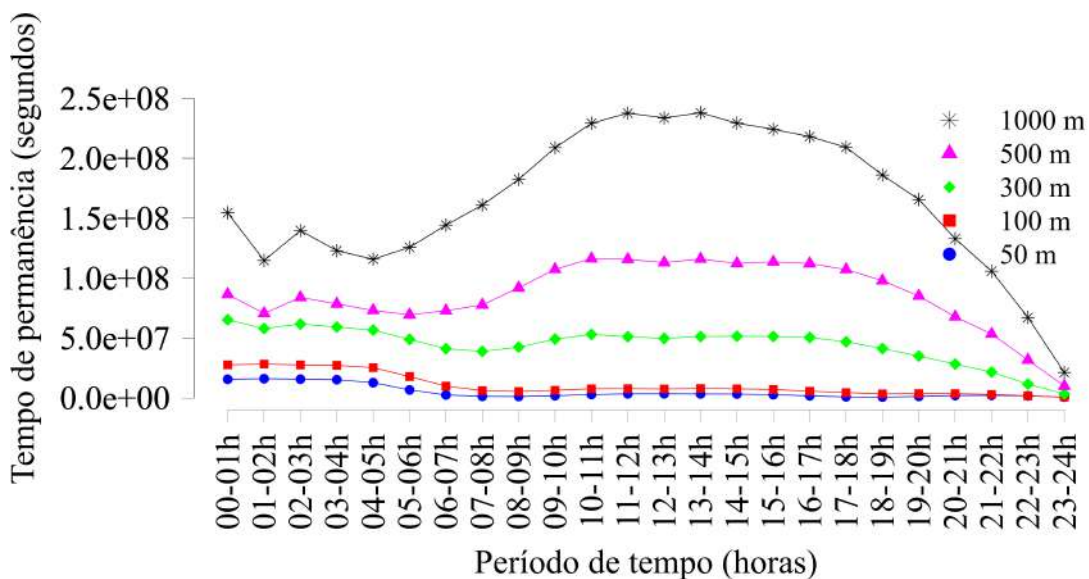


Figura 6.3: Tempo total de permanência amostrado pelo período do dia.

A Figura 6.3 apresenta o tempo de permanência em intervalos de 1 hora durante

o dia, utilizando todos os veículos e para diferentes tamanhos de clusters. Observa-se que a hora do dia influencia diretamente o tempo de permanência e conseqüentemente a capacidade da rede.

Durante o horário comercial, e especialmente nos horários de pico, há um aumento no tempo de permanência para todas as dimensões dos clusters. Nesses horários há um número maior de veículos para atender a demanda dos cidadãos se locomovendo pela cidade, onde existe um movimento entre o lar e o local de trabalho, ou estudantes indo para a escola, ou pessoas indo comprar alguma mercadoria na zona comercial da cidade, para citar alguns. O período entre as 20:00h e as 08:00h é aquele com menos tempo de permanência e, conseqüentemente, com menor capacidade de transferência de dados.

O cluster de 300 metros e, especialmente, os clusters de 50 e 100 metros apresentam também um elevado tempo de permanência nos períodos da madrugada, entre as 00:00h e as 05:00h. Esse tempo de permanência elevado é explicado pelo tempo que os ônibus permanecem na garagem. A Figura 6.4 mostra a diferença entre os ônibus e os táxis no primeiro cenário.

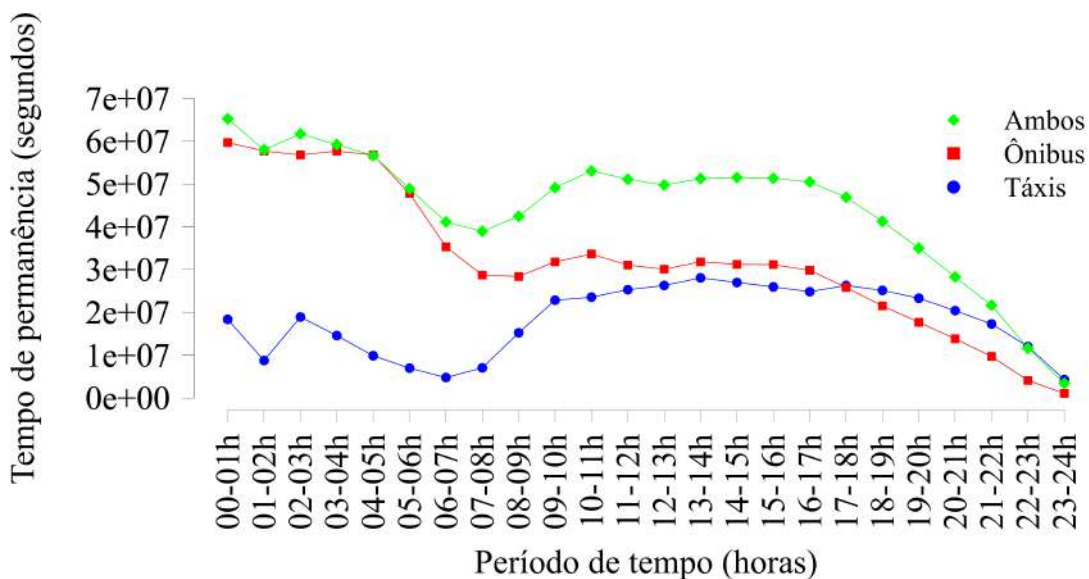


Figura 6.4: Tempo total de permanência por tipo de veículo com clusters de 300 m.

A Figura 6.4 ilustra o tempo de permanência apenas para clusters de 300 metros, diferenciado pelo tipo de veículo. A figura mostra que os táxis têm um menor tempo de permanência durante o dia, mas depois das 18:00h, eles passam a apresentar um tempo de permanência maior do que o dos ônibus. Isso é explicado principalmente porque a diminuição das atividades de ônibus é maior do que a redução da atividade táxis: à noite mais pessoas tendem a usar táxis devido ao menor número de ônibus disponíveis. O cenário em que ambos os veículos são considerados atinge um melhor rendimento em todos os casos.

6.2 Latência de envio

Para analisar a latência no envio de informações através da rede, é necessário conhecer o período que um veículo não está em contato com nenhum cluster, ou o tempo de viagem entre os clusters. Esta variável mostra o maior tempo necessário para que um veículo transmita os dados gerados e coletados no seu trajeto para um local que tenha acesso à Internet, ou seja, o tempo decorrido até entrar em contato com o ponto de acesso localizado dentro de cada um dos clusters. O pior cenário, ou o cenário com o tempo mais longo, acontece para os dados que são coletados no cluster de origem ou assim que o veículo sai do raio de contato com um cluster. A latência neste caso será igual ao tempo de viagem entre este cluster e o próximo cluster.

Os tempos de viagens entre clusters são usados para traçar o gráfico da Figura 6.5. Esta figura mostra a função de distribuição acumulada (CDF) do tempo que um veículo leva para entrar em contato com o próximo cluster, ou o tempo entre contatos, considerando ônibus e táxis simultaneamente. O eixo X representa o tempo em minutos entre contatos e o eixo Y exibe a fração dos dados. Cada linha representa um tamanho de cluster diferente, variando de 50 a 1.000 metros. A Figura 6.5 mostra que a medida que o tamanho do cluster aumenta, o tempo de contato entre clusters diminui. Com clusters de 50 metros, 90% dos contatos entre clusters decorrem em 180 minutos ou menos. Com clusters de 1.000 metros, 90% dos contatos acontecem em menos de 8 minutos. Em todos os casos, a CDF é longa, o que significa que alguns veículos têm uma alta latência ao transmitir os dados entre clusters. Noventa por cento dos tempos entre contatos são inferiores a 90 minutos no caso de clusters de 100 a 1.000 metros.

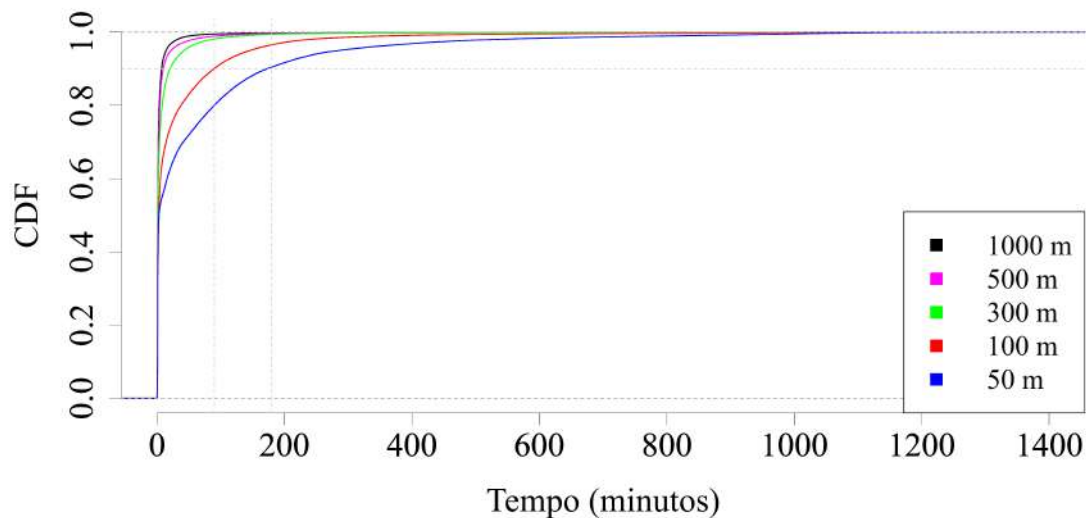


Figura 6.5: CDF do tempo que os veículos levam para alcançar o próximo cluster.

A Figura 6.6 exibe a função de distribuição acumulada (CDF) da média do tempo até chegar ao próximo cluster de cada veículo. A Figura 6.5 difere da Figura 6.6, pois a primeira é a média global de tempo de tempo de viagens, já a segunda figura representa a média de tempo que cada um dos veículos leva para chegar ao próximo cluster, onde a latência média de cada veículo é calculada.

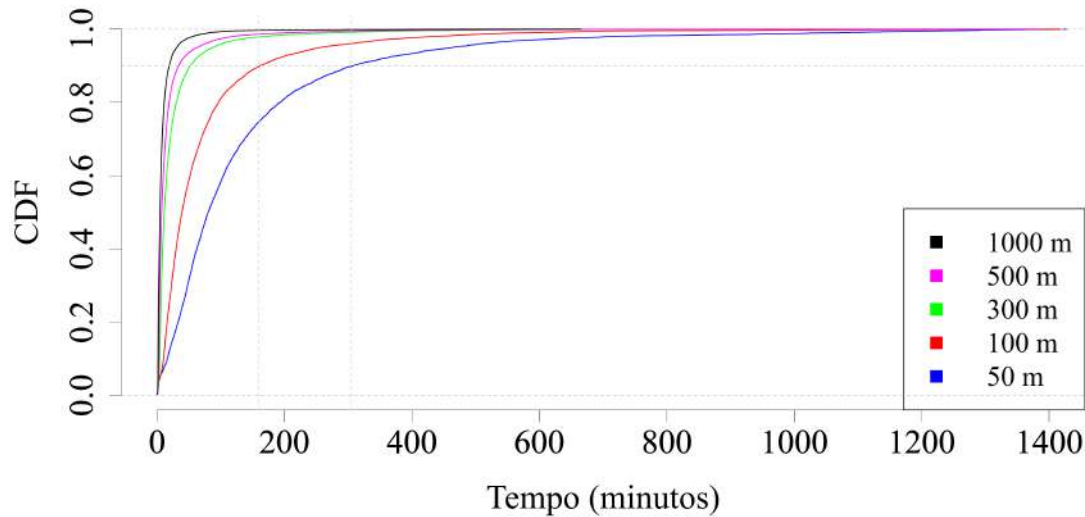


Figura 6.6: Tempo médio gasto para encontrar o próximo cluster (tempo off-line).

A Figura 6.6 representa o tempo médio que um veículo utilizou para contatar qualquer cluster ou tempo médio “off-line” de cada veículo. Em todos os cenários com tamanhos variados de cluster, a CDF tem cauda longa, o que significa que alguns veículos contatam clusters apenas algumas vezes ao dia, apresentando um elevado tempo de interconexão, ou tempo off-line. Um protocolo de roteamento deve usar essa informação para evitar a escolha dos veículos que contatam clusters poucas vezes ao dia, dando preferência a veículos com menor tempo off-line e, no caso de não possuir alternativa, enviar dados através dos veículos com maior tempo off-line, já sabendo que a latência de transmissão será alta.

Os veículos no cenário com clusters de 1.000 metros apresentam um tempo médio máximo de cerca de 11 horas em um intervalo de 24 horas. Nos clusters com 300 metros ou mais, 90% do tempo off-line é inferior a 50 minutos, atingindo 20 minutos para os clusters de 1.000 metros. Isso indica que o tempo de descarga de dados varia de acordo com o tamanho do cluster, com a possibilidade de ter um tempo médio menor do que 30 minutos se os clusters são suficientemente grandes.

A métrica denominada previsibilidade representa a chance de um veículo retornar a determinado cluster. Para medir a previsibilidade dos encontros de veículos com clusters, mede-se o número de vezes que um veículo visita um cluster que já foi visitado por ele anteriormente, dividido pelo número total de clusters que o veículo visitou em um dia. Em outras palavras, divide-se o número de clusters únicos

visitados pelo número total de clusters visitados por um veículo. Essa variável diz se o veículo tende a visitar o mesmo cluster ao longo do dia. Na Figura 6.7, agrupam-se os veículos de acordo com a sua previsibilidade medida. Valores próximos de 0 significam que todos os clusters visitados são únicos, enquanto valores próximos de 1 significam que o veículo visitou várias vezes os mesmos clusters, com poucos clusters únicos.

Na Figura 6.7, notamos que a previsibilidade cresce a medida que o tamanho do cluster cresce. O número de clusters visitados por um veículo apenas uma vez é alto nos clusters de 50 metros, levando a uma baixa previsibilidade. Para clusters de 1.000 metros, os encontros são altamente previsíveis, já que os veículos tendem a visitar os mesmos clusters mais vezes. Embora a cidade dividida em células de 50 metros tenha um maior número destas do que a de 1.000 metros, a análise considera apenas os 100 maiores clusters e a previsibilidade só é influenciada pelo número de vezes que um veículo visita cada cluster. Por fim, com uma alta previsibilidade, é possível determinar mais precisamente o contato entre um determinado veículo e um cluster, planejando melhor as transferências de dados.

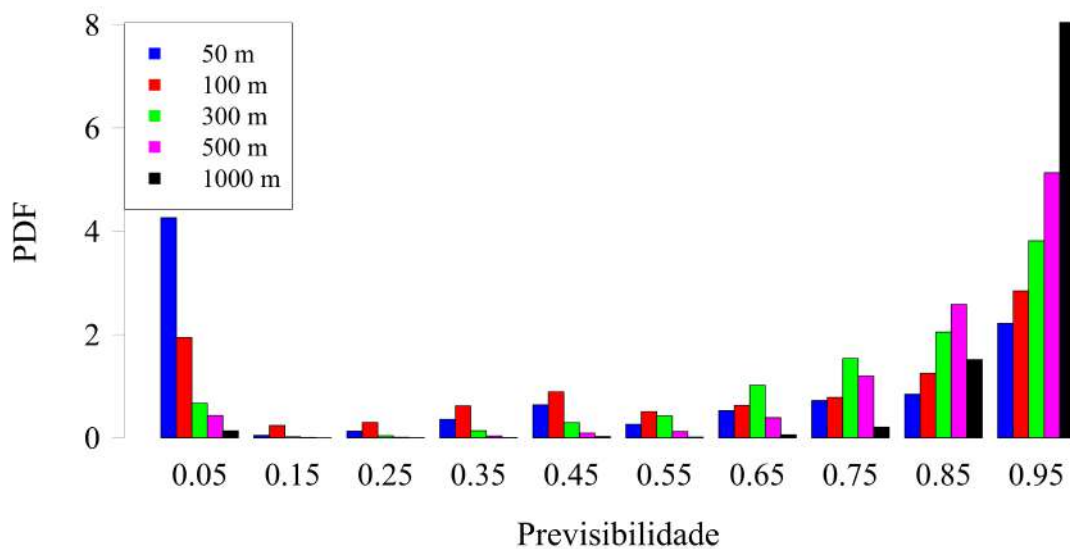


Figura 6.7: Previsibilidade de encontrar um cluster.

6.3 Conclusão

Este capítulo apresentou a análise da capacidade de transmissão de dados dependente da tecnologia, indicando a quantidade de dados que hipoteticamente podem circular através dos veículos da cidade do Rio de Janeiro e as diversas localidades contendo pontos de acesso espalhadas pela cidade. Essa análise foi realizada a partir do tempo de permanência nos clusters.

Este capítulo respondeu a segunda pergunta realizada na Introdução: Quanta informação é possível transmitir entre essas regiões em função do tempo?. A partir do tempo total de permanência dos veículos nos clusters é possível estimar a quantidade de dados trafegados ao multiplicá-lo por uma constante α . A quantidade total de dados transportados obtida, que varia de 15 a 500 terabytes quando utilizada uma conexão de 1 Mbps. Além disso, foi observado que o horário do dia interfere na quantidade transferida de dados, com um aumento na capacidade de transmissão de dados durante o dia, decaimento à noite e com um valor elevado de possibilidade de transferência na madrugada nas garagens dos ônibus. A partir das 18:00h os táxis passam a ter mais contatos com os clusters do que os ônibus, mostrando a elevação na utilização dos táxis no período noturno.

Também foi explorada a latência das transmissões de dados entre os clusters, utilizando métricas de tempo de viagens entre clusters. A latência é capaz de indicar o atraso esperado entre as transmissões de dados para cada tipo de tecnologia.

Dessa forma, nesta tese, focamos na capacidade de dados que podem ser transportados pela rede veicular e não como esses dados são gerados ou aplicados. A partir dos resultados observados é possível identificar a capacidade da rede de transportar grandes quantidades de dados entre diversos locais da cidade durante um dia.

Além disso, as ferramentas de análise propostas podem ser utilizadas com parâmetros de acordo com a necessidade, ou seja, é possível aplicar os parâmetros de uma tecnologia de transmissão conhecida e desejada para obter as características de transmissão desta rede veicular. Essa análise, por exemplo, pode ser realizada para identificar se uma determinada aplicação pode ser implementada em uma rede veicular com determinada tecnologia de transmissão, identificando de forma precoce os requisitos de transmissão da rede veicular, o número de pontos de acesso necessários e a quantidade de ônibus necessários para suprir determinado pre-requisito de uma aplicação.

Capítulo 7

Conclusões e Trabalhos Futuros

Esta tese utilizou a ferramenta de clusterização para analisar a capacidade de transporte de dados por uma rede veicular formada por ônibus e táxis que circulam na cidade do Rio de Janeiro. Com base em dados de localização reais obtidos através dos GPS instalados nesses veículos, foi idealizado um cenário composto por múltiplos pontos de coleta de dados espalhados pela cidade contendo equipamentos para comunicação por rádio e por veículos com equipamentos capazes de se comunicar com os pontos de coleta. Nesse cenário foi possível identificar o padrão de movimentação dos veículos, a quantidade possível de transferência de dados e as características temporais das transmissões, mostrando a possibilidade de transferir uma grande quantidade de dados por meio dos veículos.

Inicialmente, os dados de localização dos ônibus e táxis não estão disponíveis de forma direta. Os dados dos ônibus são disponibilizados pela prefeitura da cidade do Rio de Janeiro somente em tempo real, sendo necessária a sua coleta e armazenamento para obter dados de deslocamentos diários. Os dados dos táxis foram gentilmente cedidos pela empresa 99, através de contato com o Sr. Guilherme Mori. Ambos os dados coletados apresentam apenas intervalos da localização dos veículos na cidade, não apresentando o sincronismo e a quantidade de informações necessárias, como por exemplo a localização a cada segundo, indispensáveis para obter a exata localização dos veículos capaz de ser utilizada diretamente na análise desses dados. Os dados coletados em si apresentam falhas, como locais onde não há possibilidade da circulação de veículos, como no oceano, e períodos de tempo em que o veículo não indica a sua posição. Por isso foi necessário buscar métodos de exploração dos dados em que essas falhas não interferissem no resultado da análise realizada.

Os dados dos ônibus foram coletados, filtrados e estão disponíveis na página do Grupo de Teleinformática e Automação, além de estarem disponíveis no repositório “Crawdad” [69, 70] da Universidade de Dartmouth, dos Estados Unidos.

A metodologia para a análise inicialmente se baseou na identificação da pos-

sibilidade de usar todas as informações de localização dos veículos, ou “posições” e, conseqüentemente, os encontros entre veículos para identificar todas as possíveis trocas de mensagens entre veículos. Esse cenário se mostrou complexo demais para ser analisado em tempo hábil, precisando de uma quantidade proibitiva de recursos computacionais. Dessa forma, foi necessário a pesquisa e o desenvolvimento de algoritmos capazes de lidar com essa grande quantidade de dados de localização de veículos em tempo hábil.

Portanto, ao analisar as possibilidades de redes para a transferência de dados por veículos, foi idealizada uma topologia e foram determinadas as posições na cidade de equipamentos de rádio capazes de se comunicar com os veículos que estiverem ao seu alcance. O problema se modificou para descobrir os locais desses pontos de acesso e, a partir deles e das localizações dos veículos, identificar as características da rede, principalmente as relacionadas à transmissão de dados.

Os pontos de acesso são finitos e apresentam raios de alcance e taxas de transferência de dados que dependem da tecnologia de rede sem-fio utilizada. Nesta tese foi utilizado um algoritmo de clusterização por grades presente na literatura, o STING [65], para determinar os locais dos pontos de acesso. O algoritmo determina este local através da quantidade de veículos que indicaram estar em determinado local da grade ou região da cidade. A heurística de posicionamento compreende identificar os locais com maior quantidade são selecionados como pontos de acesso, também chamados de clusters nesta tese. O algoritmo STING, dentre os algoritmos de clusterização por grades, é o de menor complexidade, sendo indicado para o uso quando a base de dados é de grandes proporções. O algoritmo é capaz de reduzir drasticamente o tempo de análise sem deixar de usar todos os veículos da cidade (quando utilizados clusters quadrados de 1.000 metros de lado), passando do tempo de execução previsto de mais de 3.000 dias para 14 horas, no pior cenário.

De forma a não se ater a uma única tecnologia e poder analisar diversas tecnologias, a variável de alcance entre os aparelhos de rádio do ônibus e do ponto de acesso foi definida com 5 valores distintos – 50, 100, 300, 500 e 1.000 metros –, sendo dessa forma possível analisar diversas tecnologias de transmissão. Além do raio variável, não é definida a capacidade de transmissão das tecnologias, sendo utilizada uma variável α capaz de ser substituída por um valor de transmissão de dados de determinada tecnologia para obter a capacidade de transmissão de dados do sistema como um todo.

Após a identificação dos clusters, foi desenvolvido um algoritmo de cálculo de viagens, capaz de identificar e medir o tempo de todos os deslocamentos de veículos entre os clusters. Este algoritmo é usado para identificar o número de viagens que um veículo realizou entre um par de clusters, contabilizando a quantidade e o tempo de cada viagem, além do tempo que o ônibus permaneceu em contato com o ponto

de acesso. A partir desses dados são realizadas as análises que visam responder as perguntas realizadas na introdução: *Como as regiões da cidade estão interligadas em relação ao deslocamento dos veículos?* e *Quanta informação é possível transmitir entre essas regiões em função do tempo?*

Com o intuito de responder a primeira pergunta analisou-se a dinâmica da rede observando, entre outros fatos, que táxis e ônibus têm padrão de movimentação distinto, o primeiro com rotas do tipo origem-destino e o segundo seguindo rotas predefinidas por grandes vias. Além disso, os táxis têm a sua maior concentração nas áreas de maior densidade demográfica da cidade. Ônibus, na forma analisada, realizam viagens mais longas do que táxis, chegando a ter o dobro da distância máxima. A análise também mostra que as garagens dos ônibus são um local com alto tempo de permanência, podendo ser utilizadas em benefício das transferências de dados capazes de suportar elevadas latências.

Já a segunda pergunta é respondida através da análise dos tempos de permanência nos clusters e do tempo de viagens entre clusters. A partir do tempo total de permanência dos veículos nos clusters é possível estimar a quantidade de dados trafegados ao multiplicá-lo por uma constante α . A quantidade total de dados transportados obtida, que varia de 15 a 500 terabytes quando utilizada uma conexão de 1 Mbps, pode ser considerada menor do que a capacidade real de transmissão, já que as tecnologias de transmissão sem fio aplicadas a redes veiculares atingem a centenas de megabits por segundo, além dessas tecnologias de transmissão estarem em constante evolução.

Além disso, foi observado que o horário do dia interfere na quantidade transferida de dados, com um aumento na capacidade de transmissão de dados durante o dia, decaimento à noite e com um valor elevado de possibilidade de transferência na madrugada nas garagens dos ônibus. A partir das 18:00h os táxis passam a ter mais contatos com os clusters do que os ônibus, mostrando a elevação na utilização dos táxis no período noturno.

Finalmente, foi realizada uma análise para determinar o fluxo de dados na rede, onde o tempo de viagens é utilizado para descobrir o tempo médio que um veículo não está em contato com um cluster. Alguns veículos encontram clusters poucas vezes ao dia, porém, mais de 90% dos veículos têm um tempo de contato entre clusters baixo, entre 8 e 180 minutos, dependendo da tecnologia utilizada. Isso indica que o tempo médio para encontrar o próximo cluster varia de acordo com o raio de alcance da tecnologia ou tamanho do cluster, com a possibilidade de ter um tempo médio de menos de 30 minutos se os clusters forem suficientemente grandes. Ademais, a chance de que um veículo volte a um cluster aumenta de acordo com o valor do alcance, onde grandes alcances fazem com que os veículos tenham a tendência de percorrer sempre os mesmos clusters.

Com isso, pode-se dividir as contribuições desta tese em 3 grandes tópicos:

- coleta e análise dos dados de localização dos ônibus da cidade do Rio de Janeiro [71];
- elaboração de uma metodologia para análise de grande quantidade de dados de localização de veículos [72]; e
- modelagem de uma rede de transmissão de dados e análise de sua capacidade.

Além dos artigos mencionados, a partir deste trabalho foi submetido um artigo para a revista “*IEEE Transactions on Intelligent Transportation Systems*” abrangendo todas as 3 contribuições desta tese, principalmente a modelagem de uma rede de transmissão de dados, analisando a sua capacidade de transmissão.

Como trabalhos futuros há a possibilidade de utilizar dados reais de transferência de dados entre veículos para preencher de forma fidedigna as variáveis de alcance da tecnologia e capacidade de transmissão de dados (α). O desenvolvimento e instalação de equipamentos de transmissão de dados sem-fio nos veículos e em locais específicos da cidade é desejável para obter a capacidade real de transmissão de dados. O estudo com diversas tecnologias sem-fio instaladas simultaneamente nos veículos é capaz de prover comparações para melhor modelar o método proposto. Ao obter um método mais fidedigno, a partir de dados reais, estudos de comparação de capacidade serão mais precisos para a escolha da tecnologia de transmissão de dados.

Outros tipos de análises podem ser realizadas, incluindo uma análise com um período temporal ampliado, como uma semana ou um mês, obtendo métricas de variância e confiança dos dados em relação a mudança dos dias de análise. Outras métricas também podem ser exploradas, principalmente a localização dos clusters, em que outro critério de seleção das localizações dos clusters pode ser utilizado para se obter resultados distintos, como forçar a localização dos clusters em locais distantes do centro da cidade para melhorar a comunicação nessas áreas, ou variar o tamanho dos clusters para obter uma melhor distribuição dos mesmos.

Há ainda a possibilidade de criar um protocolo para o sistema de comunicação desenvolvido nesta tese com as localizações dos veículos e clusters para determinar a real capacidade de troca de mensagens. Este protocolo deve aproveitar todas as características encontradas da rede, como diferenciação de características de ônibus e táxis, diferença de tempo de contato e de tempo de viagens entre clusters, métricas de latência e previsibilidade e métricas de redes complexas, como centralidade de um vértice (cluster) e grau dos vértices. Um protocolo de troca de mensagens específico pode ser capaz de priorizar determinadas mensagens ou impedir que as mensagens se percam ao utilizar informações dos veículos, como linha do ônibus ou tendência daquele veículo voltar a determinado cluster (previsibilidade). Essa proposta cria

diversos tipos de mensagens que podem ter as suas características de transmissão e entrega diversificadas de acordo com algum fator desejado, como tempo de envio, quantidade de dados a ser enviada ou até mesmo custo.

Também é possível determinar quais dos pontos de acesso precisam estar conectados com a Internet e qual deve ser a sua banda de conexão a Internet. A característica de grande quantidade de viagens entre clusters próximos pode ser utilizada para reduzir a necessidade de conexão de certos clusters rede central, ou a Internet. Clusters próximos podem trocar uma quantidade grande de dados pois têm uma grande quantidade de veículos que circulam entre eles, dessa forma, se a rede for capaz de transportar uma quantidade maior de dados menor do que é necessário para o seu funcionamento, alguns nós podem não precisar de conexão com a Internet, encaminhando os dados para os clusters mais próximos que tenham conexão com a Internet.

Outro rumo é utilizar essas análises na fase de identificação de requisitos de uma nova aplicação para apontar possíveis cenários em que essa rede pode ser utilizada. Se um aplicativo é capaz de suportar o atraso da rede, no caso de uma comunicação que não seja em tempo real, esse tipo de rede se torna um potencial candidato. Conhecendo os requisitos da aplicação e a capacidade da tecnologia, também é possível inferir o número de clusters e número de veículos que precisam colaborar para a rede ser capaz de suportar essa aplicação. Aplicações de cidade inteligente, como sensoriamento remoto e IoT, aplicações de transferência de informações para os veículos e de transferências de grandes massas de dados entre locais da cidade, entre outras aplicações, podem se beneficiar de uma rede formada por veículos públicos urbanos e alguns pontos de acesso espalhados pela cidade.

Referências Bibliográficas

- [1] TIEN, J. M. “Big Data: Unleashing information”, *Journal of Systems Science and Systems Engineering*, v. 22, n. 2, pp. 127–151, jun. 2013.
- [2] COSTA, L. H. M. K., DE AMORIM, M. D., CAMPISTA, M. E. M., et al. “Grandes Massas de Dados na Nuvem: Desafios e Técnicas para Inovação”, *Minicursos do Simpósio Brasileiro de Redes de Computadores - SBRC'2012*, pp. 2–58, maio 2012.
- [3] ALBINO, V., BERARDI, U., DANGELICO, R. M. “Smart Cities: Definitions, Dimensions, Performance, and Initiatives”, *Journal of Urban Technology*, v. 22, n. 1, pp. 3–21, fev. 2015.
- [4] DOS S. ALVES, R., DO V. CAMPBELL, I., DE S. COUTO, R., et al. “Redes Veiculares: Princípios, Aplicações e Desafios”, *Minicursos do Simpósio Brasileiro de Redes de Computadores - SBRC'2009*, p. 199–254, maio 2009.
- [5] CAMPOLO, C., MOLINARO, A., SCOPIGNO, R. “From today’s VANETs to tomorrow’s planning and the bets for the day after”, *Vehicular Communications*, v. 2, n. 3, pp. 158–171, jul. 2015.
- [6] FETRANSPOR. “Dados operacionais mensais do Município do Rio de Janeiro – Ano de 2015”. Internet, 2017. Disponível em: http://www.fetranspordocs.com.br/downloads/TB4_DadosoperacionaismensaisdomunicipiodoRio_anode2016.xls.
- [7] LLORET, J., CANOVAS, A., CATALÁ, A., et al. “Group-based protocol and mobility model for VANETs to offer internet access”, *Journal of Network and Computer Applications*, v. 36, n. 3, pp. 1027–1038, maio 2013.
- [8] FESTAG, A. “Cooperative intelligent transport systems standards in europe”, *IEEE Communications Magazine*, v. 52, n. 12, pp. 166–172, dez. 2014.

- [9] MITTON, N., RIVANO, H. “On the use of city bikes to make the city even smarter”. In: *International Conference on Smart Computing Workshops*, pp. 3–8, nov. 2014.
- [10] GORCITZ, R. A., JARMA, Y., SPATHIS, P., et al. “Vehicular carriers for big data transfers”. In: *IEEE Vehicular Networking Conference (VNC)*, pp. 109–114, nov. 2012.
- [11] LAOUTARIS, N., SIRIVIANOS, M., YANG, X., et al. “Inter-datacenter bulk transfers with netstitcher”. In: *ACM SIGCOMM Computer Communication Review*, v. 41, pp. 74–85. ACM, ago. 2011.
- [12] KERÄNEN, A., OTT, J. “DTN over aerial carriers”. In: *4th ACM Workshop on Challenged Networks*, pp. 67–76. ACM, set. 2009.
- [13] ZARAFSHAN-ARAKI, M., CHIN, K.-W. “TrainNet: A transport system for delivering non real-time data”, *Computer Communications*, v. 33, n. 15, pp. 1850–1863, set. 2010.
- [14] DOS SANTOS, A., DE LIMA BRAGA, M., VELLOSO, P. B., et al. “Análise da Capacidade de uma Rede Tolerante a Atrasos e Desconexões na Baía Hidrográfica Amazonense”, *XXXII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos - SBRC*, p. 14, maio 2014.
- [15] MAZLOUMI, E., CURRIE, G., ROSE, G. “Using GPS data to gain insight into public transport travel time variability”, *Journal of Transportation Engineering*, v. 136, n. 7, pp. 623–631, jul. 2009.
- [16] SOROUSH, H., BANERJEE, N., BALASUBRAMANIAN, A., et al. “DOME: a diverse outdoor mobile testbed”. In: *1st ACM International Workshop on Hot Topics of Planet-Scale Mobility Measurements*, p. 2. ACM, jun. 2009.
- [17] ZHOU, C., DAI, P., WANG, F., et al. “Predicting the passenger demand on bus services for mobile users”, *Pervasive and Mobile Computing*, jan. 2015.
- [18] SEDE, M., LI, X., LI, D., et al. “Routing in Large-Scale Buses Ad Hoc Networks”. In: *IEEE Wireless Communications and Networking Conference*, pp. 2711–2716, mar. 2008.
- [19] DOERING, M., PÖTTNER, W.-B., PÖGEL, T., et al. “Impact of radio range on contact characteristics in bus-based delay tolerant networks”. In: *Wireless On-Demand Network Systems and Services (WONS)*, pp. 195–202. IEEE, jan. 2011.

- [20] DOERING, M., PÖGEL, T., PÖTTNER, W.-B., et al. “A New Mobility Trace for Realistic Large-scale Simulation of Bus-based DTNs”. In: *ACM Workshop on Challenged Networks*, pp. 71–74, set. 2010.
- [21] DOERING, M., PÖGEL, T., WOLF, L. “DTN routing in urban public transport systems”. In: *5th ACM workshop on Challenged networks*, pp. 55–62. ACM, set. 2010.
- [22] DOERING, M., WOLF, L. “Opportunistic vehicular networking: Large-scale bus movement traces as base for network analysis”. In: *High Performance Computing & Simulation (HPCS)*, pp. 671–678. IEEE, jul. 2015.
- [23] JIANG, X., DU, D. H. “BUS-VANET: A BUS vehicular network integrated with traffic infrastructure”, *Intelligent Transportation Systems Magazine*, v. 7, n. 2, pp. 47–57, abr. 2015.
- [24] NABOULSI, D., FIORE, M. “On the instantaneous topology of a large-scale urban vehicular network: the Cologne case”. In: *Fourteenth ACM international symposium on Mobile ad hoc networking and computing*, pp. 167–176. ACM, jul. 2013.
- [25] ARANITI, G., CAMPOLO, C., CONDOLUCI, M., et al. “LTE for vehicular networking: a survey”, *IEEE Communications Magazine*, v. 51, n. 5, pp. 148–157, maio 2013.
- [26] PIRO, G., ORSINO, A., CAMPOLO, C., et al. “D2D in LTE vehicular networking: System model and upper bound performance”. In: *7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 281–286, out. 2015.
- [27] LOCHERT, C., HARTENSTEIN, H., TIAN, J., et al. “A routing strategy for vehicular ad hoc networks in city environments”. In: *IEEE Intelligent Vehicles Symposium*, pp. 156–161, jun. 2003.
- [28] BAI, F., KRISHNAMACHARI, B. “Spatio-temporal Variations of Vehicle Traffic in VANETs: Facts and Implications”. In: *Sixth ACM International Workshop on VehiculAr InterNETworking*, pp. 43–52. ACM, set. 2009.
- [29] UPPOOR, S., TRULLOLS-CRUCES, O., FIORE, M., et al. “Generation and Analysis of a Large-Scale Urban Vehicular Mobility Dataset”, *IEEE Transactions on Mobile Computing*, v. 13, n. 5, pp. 1061–1075, maio 2014.

- [30] JETCHEVA, J. G., HU, Y. C., PALCHAUDHURI, S., et al. “Design and evaluation of a metropolitan area multitier wireless ad hoc network architecture”. In: *Mobile Computing Systems and Applications*, pp. 32–43, out. 2003.
- [31] BURGESS, J., GALLAGHER, B., JENSEN, D., et al. “MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks”. In: *IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1–11, abr. 2006.
- [32] HOQUE, M. A., HONG, X., DIXON, B. “Analysis of mobility patterns for urban taxi cabs”. In: *International Conference on Computing, Networking and Communications (ICNC)*, pp. 756–760, jan. 2012.
- [33] BEJAN, A. I., GIBBENS, R. J., EVANS, D., et al. “Statistical modelling and analysis of sparse bus probe data in urban areas”. In: *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1256–1263. IEEE, set. 2010.
- [34] ZHANG, L., JIA, S., LIU, Z., et al. “Bus-Ads: Bus Trajectory-Based Advertisement Distribution in VANETs Using Coalition Formation Games”, *IEEE Systems Journal*, v. PP, n. 99, pp. 1–1, set. 2015.
- [35] SCHILDT, S., WOLF, L. “Goodies for data: Game-based data propagation in DTNs”. In: *IEEE Globecom Workshops*, pp. 320–325. IEEE, dez. 2012.
- [36] LI, H., WANG, B., SONG, Y., et al. “VeShare: a D2D infrastructure for real-time social-enabled vehicle networks”, *IEEE Wireless Communications*, v. 23, n. 4, pp. 96–102, ago. 2016.
- [37] BARON, B., CAMPISTA, M., SPATHIS, P., et al. “Virtualizing vehicular node resources: Feasibility study of virtual machine migration”, *Vehicular Communications*, v. 4, n. Supplement C, pp. 39–46, abr. 2016.
- [38] MEZGHANI, F., DHAOU, R., NOGUEIRA, M., et al. “Content dissemination in vehicular social networks: taxonomy and user satisfaction”, *IEEE Communications Magazine*, v. 52, n. 12, pp. 34–40, dez. 2014.
- [39] KOSAR, T., ARSLAN, E., ROSS, B., et al. “Storkcloud: Data transfer scheduling and optimization as a service”. In: *4th ACM workshop on Scientific cloud computing*, pp. 29–36. ACM, jun. 2013.

- [40] CHO, B., GUPTA, I. “Budget-constrained bulk data transfer via internet and shipping networks”. In: *8th ACM international conference on Autonomic computing*, pp. 71–80. ACM, jun. 2011.
- [41] LAOUTARIS, N., SMARAGDAKIS, G., RODRIGUEZ, P., et al. “Delay tolerant bulk data transfers on the Internet”. In: *ACM SIGMETRICS Performance Evaluation Review*, v. 37, pp. 229–238. ACM, jun. 2009.
- [42] WAN, J., ZHANG, D., ZHAO, S., et al. “Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions”, *IEEE Communications Magazine*, v. 52, n. 8, pp. 106–113, ago. 2014.
- [43] YAMAMURA, S., NAGATA, A., TSURU, M. “Store-carry-forward based networking infrastructure: Vision and potential”. In: *Intelligent Networking and Collaborative Systems (INCoS)*, pp. 594–599. IEEE, nov. 2011.
- [44] YAMAMURA, S., NAGATA, A., TSURU, M., et al. “Virtual segment: Store-carry-forward relay-based support for wide-area non-real-time data exchange”, *Simulation Modelling Practice and Theory*, v. 19, n. 1, pp. 30–46, jan. 2011.
- [45] IBGE. “Cidades@ - Rio de Janeiro”. Internet, 2017. Disponível em: <<http://cidades.ibge.gov.br/xtras/perfil.php?codmun=3304554>>.
- [46] ARMAZÉM DE DADOS. “Rio em síntese”. Internet, 2017. Disponível em: <<http://www.armazemdedados.rio.rj.gov.br/>>.
- [47] SECRETARIA MUNICIPAL DE TRANSPORTES - SMTR. “PERGUNTAS FREQUENTES”. Internet, 2017. Disponível em: <<http://www.rio.rj.gov.br/web/smtr/exibeconteudo?id=6254448>>.
- [48] FETRANSPOR. “Mobilidade em Números”. Internet, 2017. Disponível em: <<https://www.fetranspor.com.br/mobilidade-urbana-setor-em-numeros>>.
- [49] 99. “About us”. Website, nov. 2017. Disponível em: <<http://www.99taxi.com/en/startup/>>.
- [50] DATA.RIO. “Transporte e mobilidade da cidade do Rio de Janeiro”. Internet, 2016. Disponível em: <<http://data.rio/organization/0a61fd92-731e-4bd6-ae08-ba8bf5093f0c?groups=transporte-e-mobilidade>>.

- [51] DATA.RIO. “Pontos de parada das linhas do ônibus”. Internet, 2016. Disponível em: <<http://data.rio/dataset/pontos-de-parada-de-onibus>>.
- [52] DATA.RIO. “Pontos dos trajetos das linhas de ônibus”. Internet, 2016. Disponível em: <<http://data.rio/dataset/pontos-dos-percursos-de-onibus>>.
- [53] DATA.RIO. “GPS dos ônibus”. Internet, 2016. Disponível em: <<http://data.rio//dataset/gps-de-onibus>>.
- [54] PREFEITURA DA CIDADE DO RIO DE JANEIRO. “IplanRio - Empresa Municipal de Informática”. Internet, 2016. Disponível em: <<http://www.rio.rj.gov.br/web/iplanrio>>.
- [55] DATA.RIO. “Documentação Posições GPS dos ônibus”. Internet, 2016. Disponível em: <http://dadosabertos.rio.rj.gov.br/apitransporte/apresentacao/pdf/documentacao_gps.pdf>.
- [56] PREFEITURA DA CIDADE DO RIO DE JANEIRO. “Relatório Diário de Operação - RDO das Linhas”. Internet, 2017. Disponível em: <http://www.rio.rj.gov.br/dlstatic/10112/5026693/4132413/RDO_outubro.pdf>.
- [57] CONGRESSO NACIONAL. “Código de Trânsito Brasileiro - art. 61”. Internet, 1997. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L9503.htm>.
- [58] RIO ÔNIBUS. “BRs em funcionamento na cidade do Rio”. Internet, 2017. Disponível em: <<https://rjbr.org/r/rio-de-janeiro/transportes/brs/o-brs/>>.
- [59] FERNANDO CAVALLIERI, G. P. L. “Índice de Desenvolvimento Social - IDS: comparando as realidades microurbanas da cidade do Rio de Janeiro”. Internet, 2008. Disponível em: <http://portalgeo.rio.rj.gov.br/estudoscariocas/download/2394_%C3%8Dndice%20de%20Desenvolvimento%20Social_IDS.pdf>.
- [60] LIN, M., HSU, W.-J. “Mining GPS data for mobility patterns: A survey”, *Pervasive and Mobile Computing*, v. 12, n. Supplement C, pp. 1–16, dez. 2014.

- [61] KOPERSKI, K., ADHIKARY, J., HAN, J. “Spatial Data Mining: Progress and Challenges - Survey paper”. In: *SIGMOD Workshop on research issues on data mining and knowledge discovery*, pp. 1–10, jun. 1996.
- [62] RODDICK, J., LEES, B. G. “Paradigms for spatial and spatio-temporal data mining”, *Geographic Data Mining and Knowledge Discovery*, pp. 33–49, 2001.
- [63] RODDICK, J. F., LEES, B. G. “Spatio-temporal data mining paradigms and methodologies”. In: Miller, H. J., Han, J. (Eds.), *Geographic Data Mining and Knowledge Discovery*, cap. 2, pp. 27–44, Chapman & Hall/CRC, 2009.
- [64] KOLATCH, E. “Clustering Algorithms for Spatial Databases: A Survey”. Internet, 2001. Disponível em: <<https://pdfs.semanticscholar.org/4db2/a23114110c0d9c586c02737508f3ec71ee26.pdf>>.
- [65] WANG, W., YANG, J., MUNTZ, R., et al. “STING: A statistical information grid approach to spatial data mining”. In: *23rd International Conference on Very Large Data Bases*, pp. 186–195, ago. 1997.
- [66] DHILLON, H. S., HUANG, H., VISWANATHAN, H. “Wide-area Wireless Communication Challenges for the Internet of Things”, *IEEE Communications Magazine*, v. 55, n. 2, pp. 168–174, fev. 2017.
- [67] LIU, J., KATO, N., MA, J., et al. “Device-to-Device Communication in LTE-Advanced Networks: A Survey”, *IEEE Communications Surveys Tutorials*, v. 17, n. 4, pp. 1923–1940, dez. 2014.
- [68] ARMAZÉM DE DADOS. “Densidade demográfica, por Bairros - 2000”. Internet, 2017. Disponível em: <<http://www.data.rio/datasets/53d51e2ab0304ed38dd9bc85fd6c6b1b>>.
- [69] KOTZ, D., HENDERSON, T., ABYZOV, I., et al. “CRAWDAD dataset dartmouth/campus (v. 2009-09-09)”. Internet, set. 2009. Disponível em: <<https://crawdad.org/dartmouth/campus/20090909>>.
- [70] DIAS, D., COSTA, L. H. M. K. “CRAWDAD dataset coppe-ufrj/RioBuses (v. 2018-02-20)”. Website, fev. 2018. Disponível em: <<https://crawdad.org/coppe-ufrj/RioBuses/20180220>>.
- [71] DIAS, D. S., COSTA, L. H. M. K. “Análise da Capacidade de Dados de uma Rede de Ônibus Urbanos”, *XXXIV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, ago. 2016.

- [72] DIAS, D. S., COSTA, L. H. M. K., DE AMORIM, M. D. “Capacity analysis of a city-wide V2V network”, *7th International Conference on the Network of the Future (NOF)*, pp. 1–3, nov. 2016.

Apêndice A

Tabelas com os resultados da análise preliminar

A.1 Dados de Contatos de um Ônibus

A Seção 4.1.1 apresenta a análise preliminar dos dados de localização dos ônibus da cidade do Rio de Janeiro. Os dados obtidos são exibidos em diversas tabelas, e, por questões de melhor compreensão, não são apresentadas e analisados todos os dados na Seção 4.1.1.2.

Dessa forma, este apêndice tem a função de apresentar todas as tabelas com todos os resultados obtidos a partir dos dados da análise preliminar. Além disso, os resultados são comentados, evidenciando as suas principais características.

A.1.1 Quantidade de Contatos

Para a análise, a métrica principal a ser estudada é a quantidade de contatos que um ônibus teve durante um dia. Dentro da janela de tempo e do raio analisados, durante o período de 24h, os ônibus têm a quantidade de contatos descrita na Tabela A.1.

Apesar das linhas analisadas serem de áreas distintas da cidade e serem controladas por empresas diferentes, os dados de contatos são semelhantes, tendo diferenças menores do que o triplo do menor valor quando são considerados raios de 50 metros e de menos do que o dobro quando são considerados raios de 300 metros.

A.1.2 Quantidade de ônibus únicos contactados

Outro dado importante é a quantidade de ônibus únicos que são contactados em 24h, exibida na Tabela A.2. Como esperado, a quantidade de ônibus únicos cresce de acordo com o raio e com a janela de tempo. A variação entre o raio de 50 metros

Tabela A.1: Quantidade de contatos.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	3714	3252	3426	3023	2446	3010	2638	2062	3254	27976
100	1	7759	6055	6111	5487	4165	6010	4844	3406	8058	55303
200	1	12882	11773	12417	8710	8714	11986	9739	6284	18298	94722
300	1	16408	18035	19520	11542	14598	17731	16015	10534	23683	159138
50	2	6615	5710	5655	5152	3928	5071	4400	3286	5710	54636
100	2	14734	11328	10944	10160	7378	11030	8740	5975	15143	109087
200	2	25008	22843	23594	16667	16416	22932	18272	11765	35411	187496
300	2	32109	35325	38005	22568	28288	34385	30628	20214	46223	315793
50	5	15108	12771	12121	11331	8653	11511	9693	6952	12999	133852
100	5	35342	27050	25054	23752	17396	26328	20601	13550	36307	268729
200	5	61021	55447	56869	39931	39837	56155	44252	27866	86396	463039
300	5	78750	86335	93181	54833	69433	84505	75147	49092	112931	780686
50	10	28966	23566	22199	21305	15923	21608	18229	12733	24772	264660
100	10	69039	51850	47508	45801	33421	50503	39810	25734	70780	532604
200	10	120334	108271	110381	77919	77833	109590	87416	53979	169183	918174
300	10	155603	169426	182043	107600	136615	165286	148794	95974	221810	1549602

e a janela de 1 minuto para o raio de 300 metros e a janela de 10 minutos pode chegar a 8 vezes, no caso do ônibus D877380.

A Tabela A.3 mostra a porcentagem de ônibus únicos contactados por cada ônibus analisado, ou seja, a relação entre os ônibus únicos e os contatos totais daquele ônibus. Com uma janela de tempo pequena, a quantidade de ônibus únicos contactados é proporcionalmente grande.

Tabela A.2: Quantidade de ônibus únicos contactados.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	576	828	854	475	501	635	566	419	604	781
100	1	1184	1536	1557	938	915	1124	1173	770	1107	1265
200	1	2160	2647	2705	1775	1755	1833	2211	1402	1903	1783
300	1	2829	3391	3422	2364	2400	2112	2958	2106	2599	1965
50	2	813	1169	1177	684	679	775	854	563	756	913
100	2	1612	1975	2016	1248	1195	1293	1641	996	1344	1391
200	2	2644	3080	3150	2139	2040	1990	2702	1695	2219	1866
300	2	3294	3740	3742	2710	2642	2240	3430	2418	2917	2045
50	5	1421	1844	1875	1097	1106	1080	1406	939	1196	1190
100	5	2441	2787	2806	1808	1733	1610	2400	1519	1902	1633
200	5	3433	3752	3752	2743	2523	2241	3443	2276	2790	2012
300	5	4011	4319	4160	3267	3019	2414	4045	2948	3402	2141
50	10	2139	2560	2627	1590	1638	1393	2093	1375	1702	1411
100	10	3263	3510	3480	2440	2296	1915	3185	2071	2548	1813
200	10	4157	4398	4201	3374	2939	2486	4128	2878	3398	2117
300	10	4607	4887	4470	3835	3336	2627	4613	3520	3935	2224

A.1.3 Média de contatos

A Tabela A.4 mostra o número médio de contatos que ocorre a cada instante analisado. Esse dado indica a capacidade de conexões com outros ônibus que deter-

Tabela A.3: Porcentagem de ônibus únicos contactados.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	16%	25%	25%	16%	20%	21%	21%	20%	19%	3%
100	1	15%	25%	25%	17%	22%	19%	24%	23%	14%	2%
200	1	17%	22%	22%	20%	20%	15%	23%	22%	10%	2%
300	1	17%	19%	18%	20%	16%	12%	18%	20%	11%	1%
50	2	12%	20%	21%	13%	17%	15%	19%	17%	13%	2%
100	2	11%	17%	18%	12%	16%	12%	19%	17%	9%	1%
200	2	11%	13%	13%	13%	12%	9%	15%	14%	6%	1%
300	2	10%	11%	10%	12%	9%	7%	11%	12%	6%	1%
50	5	9%	14%	15%	10%	13%	9%	15%	14%	9%	1%
100	5	7%	10%	11%	8%	10%	6%	12%	11%	5%	1%
200	5	6%	7%	7%	7%	6%	4%	8%	8%	3%	0%
300	5	5%	5%	4%	6%	4%	3%	5%	6%	3%	0%
50	10	7%	11%	12%	7%	10%	6%	11%	11%	7%	1%
100	10	5%	7%	7%	5%	7%	4%	8%	8%	4%	0%
200	10	3%	4%	4%	4%	4%	2%	5%	5%	2%	0%
300	10	3%	3%	2%	4%	2%	2%	3%	4%	2%	0%

minado ônibus tem em uma janela de tempo. Desse número é possível identificar que ao aumentar a janela e o raio de busca, o número de ônibus encontrados é ampliado, em consequência, o número de transmissões simultâneas também é ampliado, podendo levar a uma maior interferência nas transmissões de dados. A quantidade média de contatos é bastante distinta entre os ônibus, porém o número médio de contatos sempre aumenta de acordo com o raio e de acordo com a janela de tempo, mostrando a importância da definição correta do equipamento de radiocomunicação, evitando interferências.

Tabela A.4: Média de contatos.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	4,3	3,6	3,0	3,3	2,6	3,2	2,8	2,4	4,0	20,7
100	1	8,9	6,7	5,3	6,1	4,4	6,4	5,2	3,9	9,9	41,0
200	1	14,8	13,0	10,7	9,6	9,2	12,7	10,4	7,3	22,4	70,2
300	1	18,8	20,0	16,9	12,8	15,5	18,7	17,0	12,2	29,0	117,9
50	2	7,6	6,3	4,9	5,7	4,2	5,4	4,7	3,8	7,0	40,5
100	2	16,9	12,5	9,5	11,2	7,8	11,7	9,3	6,9	18,5	80,8
200	2	28,7	25,3	20,4	18,4	17,4	24,2	19,4	13,6	43,3	138,9
300	2	36,8	39,1	32,9	24,9	29,9	36,4	32,6	23,3	56,6	233,9
50	5	17,3	14,1	10,5	12,5	9,2	12,2	10,3	8,0	15,9	99,2
100	5	40,5	29,9	21,7	26,3	18,4	27,8	21,9	15,7	44,4	199,1
200	5	69,9	61,3	49,2	44,1	42,2	59,4	47,1	32,2	105,8	343,0
300	5	90,2	95,5	80,5	60,6	73,5	89,3	79,9	56,7	138,2	578,3
50	10	33,2	26,1	19,2	23,5	16,9	22,8	19,4	14,7	30,3	196,0
100	10	79,1	57,4	41,1	50,6	35,4	53,4	42,4	29,7	86,6	394,5
200	10	137,8	119,8	95,4	86,1	82,4	115,9	93,0	62,3	207,1	680,1
300	10	178,2	187,4	157,3	118,9	144,6	174,7	158,3	110,8	271,5	1147,9

A.1.4 Número de veículos contactados no ponto de contatos máximos

O instante analisado com maior número de contatos simultâneos de ônibus em um dia é exibido na Tabela A.5. Há um aumento linear na quantidade de ônibus contactados no instante com maior número de contatos simultâneos, na maioria dos casos analisados.

O raio de alcance do ônibus, entretanto, não varia linearmente. Além disso, não há diferença significativa de número máximo de contatos ao se utilizar o raio de 200 ou o raio de 300 metros, em alguns casos tendo achado o mesmo número máximo de ônibus.

Tabela A.5: Número de veículos contactados no ponto de contatos máximos.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	41	18	18	56	16	47	31	14	34	83
100	1	108	33	40	91	25	116	56	26	114	170
200	1	130	89	90	109	32	460	82	34	209	370
300	1	130	114	113	114	48	553	92	43	244	460
50	2	47	22	27	60	19	48	33	15	36	87
100	2	119	54	58	97	29	128	63	30	121	177
200	2	133	124	129	119	41	499	92	39	219	401
300	2	133	149	168	125	76	580	121	59	258	482
50	5	50	49	51	62	26	50	41	24	42	91
100	5	125	102	97	102	39	142	89	34	137	186
200	5	140	175	182	125	74	524	136	69	240	440
300	5	168	212	222	138	127	603	177	105	279	518
50	10	62	88	92	70	41	56	66	34	52	97
100	10	153	193	167	111	70	153	137	53	143	191
200	10	255	273	282	145	138	543	215	123	258	464
300	10	297	307	323	193	203	625	277	178	299	542

A.1.5 Linhas distintas

A quantidade de linhas distintas que um ônibus encontra durante o período de tempo analisado, que é exibida na Tabela A.6. O número de linhas distintas encontrada relaciona-se a possibilidades de enviar dados para trajetórias distintas da trajetória do ônibus analisado. Quanto maior o número de encontros com linhas diferentes, maior o alcance dos dados em relação à distância. Dessa forma, um ônibus que tem encontros com uma grande quantidade de linhas, é capaz de espalhar os dados de forma mais eficaz entre as diversas regiões da cidade, precisando de menos intermediários para que o dado chegue ao destino. O número de linhas varia, principalmente, de acordo com o raio de contato entre os ônibus. A variação de linhas contactadas é pequena entre uma janela de 1 minuto e uma janela de 10 minutos.

Tabela A.6: Número de linhas únicas encontradas.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	156	194	209	129	142	89	186	108	157	70
100	1	234	252	257	202	172	113	256	157	220	81
200	1	294	301	289	269	227	136	304	209	286	90
300	1	317	322	298	294	252	146	335	248	318	93
50	2	183	225	238	158	167	101	224	131	189	76
100	2	267	276	274	230	197	121	283	179	256	83
200	2	307	311	293	279	233	140	321	229	311	90
300	2	322	331	300	301	257	150	351	265	329	95
50	5	243	265	263	208	190	118	267	168	236	84
100	5	290	304	290	263	211	132	315	205	293	87
200	5	322	328	301	300	241	153	351	259	330	93
300	5	338	344	304	313	267	160	376	285	346	100
50	10	272	285	284	239	211	130	305	198	274	89
100	10	313	326	301	285	226	147	341	229	317	93
200	10	340	345	307	313	250	164	372	282	343	97
300	10	348	358	310	326	274	171	391	305	353	102

A.1.6 Média de tempo dos contatos

A Tabela A.7 apresenta dados dos tempos médios de contato que um ônibus tem com outro ônibus. O tempo, representado na tabela em segundos, apresenta a média máxima na maioria dos casos entre 50 e 100 metros de raio. Ao aumentar a janela, mais ônibus que não estão nas mesmas vias e que não fazem caminhos semelhantes são contactados. Esses ônibus naturalmente têm uma menor oportunidade de contato, reduzindo a média de tempo de contato. A janela de contato, de forma semelhante, também amplia o número de ônibus contactados que não têm itinerário parecido com o do ônibus que está sendo analisado, fazendo com que o tempo médio de contato decresça na maioria dos casos.

Tabela A.7: Média de tempo dos contatos (segundos).

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	2555	2654	3254	2520	2991	5313	2414	3630	1059	4609
100	1	3605	3852	4539	4955	4165	5262	4482	4556	1901	5572
200	1	3720	3876	5467	4859	4184	5453	4765	4180	2467	4723
300	1	3929	3355	4622	4477	4454	2576	4155	3683	2226	4182
50	2	2315	2400	2765	1780	1939	3172	2659	1930	1296	2905
100	2	3261	2416	2925	2312	2570	2374	3153	2138	1094	2959
200	2	2136	1623	2064	1858	1857	1166	2228	1669	1283	2025
300	2	1533	1067	1347	1363	982	701	1749	1154	953	2696
50	5	3001	2474	2277	3071	1816	2584	3453	2240	1245	2812
100	5	2508	1761	1941	2323	1766	1410	2795	2251	1077	2340
200	5	1222	1160	1206	987	989	976	1813	1204	744	1795
300	5	1058	713	676	785	591	606	1114	731	637	2297
50	10	3508	2753	2733	3428	2394	2280	3803	2463	1166	3262
100	10	2375	1625	1626	1793	1582	1187	2572	1749	1021	2156
200	10	963	926	824	821	646	649	1187	1042	709	2079
300	10	755	611	487	563	294	625	688	627	607	3048

A.1.7 Contatos múltiplos

Os contatos com outros ônibus são divididos em dois tipos: Os contatos únicos são aqueles em que, para o mesmo ônibus em período correspondente a uma janela de tempo, são contactados somente uma vez; Os contatos múltiplos são aqueles que o mesmo ônibus é contactado mais de uma vez na janela de tempo. Um contato múltiplo garante que o contato entre os dois ônibus é realizado por mais tempo, fazendo com que a qualidade da conexão seja melhor, possibilitando uma maior troca de dados.

A Tabela A.8 exibe o número de contatos múltiplos. A Tabela A.9 mostra a relação entre os contatos múltiplos e os ônibus únicos contactados no banco de dados para cada um dos ônibus. Ao aumentar a janela e o raio, a proporção de contatos múltiplos aumenta, aumentando a oportunidade de cada ônibus contactado trocar dados de forma satisfatória, devido ao aumento da janela de tempo de contato.

Tabela A.8: Quantidade de contatos múltiplos.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	267	403	414	236	260	340	245	214	295	505
100	1	542	804	817	444	498	708	549	397	639	920
200	1	1247	1658	1785	970	1175	1418	1293	842	1255	1499
300	1	1885	2522	2636	1485	1847	1790	2013	1453	1802	1793
50	2	441	678	675	359	417	548	438	351	478	722
100	2	949	1333	1328	772	792	1033	944	627	932	1220
200	2	1928	2421	2543	1563	1637	1769	2001	1263	1727	1727
300	2	2627	3219	3309	2201	2322	2074	2807	1975	2445	1936
50	5	852	1230	1270	664	731	833	831	590	781	962
100	5	1713	2188	2177	1295	1306	1380	1644	1081	1442	1464
200	5	2878	3316	3415	2269	2187	2064	2839	1858	2351	1930
300	5	3587	3964	3962	2932	2819	2304	3632	2598	3068	2084
50	10	1390	1866	2000	1057	1173	1143	1345	942	1206	1213
100	10	2561	2959	3032	1885	1879	1686	2453	1627	2038	1685
200	10	3726	4036	3976	2935	2718	2331	3651	2449	2985	2066
300	10	4301	4661	4338	3533	3206	2534	4285	3148	3621	2180

A.1.8 Velocidade média de contato

Como no banco de dados está presente a informação de velocidade dos ônibus, para fins de análise, essa informação é utilizado como velocidade na hora do contato. A velocidade média de contato, exibida na Tabela A.10, é a média da velocidade dos ônibus que o ônibus analisado contacta durante um dia. Como os ônibus compartilham a mesma infraestrutura rodoviária, as condições da via estão ligadas diretamente à velocidade dos ônibus. Ainda, como o raio de contato é pequeno, ônibus próximos tendem a ter a mesma velocidade.

Aliado às condições da via, o tipo de medição das posições dos ônibus informadas ao banco de dados é determinante para a tendência de contatos em baixa

Tabela A.9: Proporção de contatos múltiplos.

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	46%	49%	48%	50%	52%	54%	43%	51%	49%	65%
100	1	46%	52%	52%	47%	54%	63%	47%	52%	58%	73%
200	1	58%	63%	66%	55%	67%	77%	58%	60%	66%	84%
300	1	67%	74%	77%	63%	77%	85%	68%	69%	69%	91%
50	2	54%	58%	57%	52%	61%	71%	51%	62%	63%	79%
100	2	59%	67%	66%	62%	66%	80%	58%	63%	69%	88%
200	2	73%	79%	81%	73%	80%	89%	74%	75%	78%	93%
300	2	80%	86%	88%	81%	88%	93%	82%	82%	84%	95%
50	5	60%	67%	68%	61%	66%	77%	59%	63%	65%	81%
100	5	70%	79%	78%	72%	75%	86%	69%	71%	76%	90%
200	5	84%	88%	91%	83%	87%	92%	82%	82%	84%	96%
300	5	89%	92%	95%	90%	93%	95%	90%	88%	90%	97%
50	10	65%	73%	76%	66%	72%	82%	64%	69%	71%	86%
100	10	78%	84%	87%	77%	82%	88%	77%	79%	80%	93%
200	10	90%	92%	95%	87%	92%	94%	88%	85%	88%	98%
300	10	93%	95%	97%	92%	96%	96%	93%	89%	92%	98%

velocidade. Como os intervalos de medições são longos, no melhor caso de 1 minuto, a probabilidade que dois ônibus em alta velocidade, ou mesmo de um parado e de outro em alta velocidade, se encontrem é menor do que dois ônibus que estejam em velocidade baixa. Dessa forma, a velocidade média global é baixa.

Tabela A.10: Velocidade média de contato (km/h).

	Linha	371	363	415	624	433	550	2336	474	353	886
	Ônibus	C516370	C825600	A481390	B516440	A275950	C302230	D877380	A291850	C475260	C302810
Raio	Janela										
50	1	7,03	8,73	12,59	7,84	10,79	8,41	14,43	12,26	5,78	1,05
100	1	5,82	8,54	11,53	6,85	10,58	6,09	12,05	11,26	3,92	0,98
200	1	6,96	8,83	11,74	8,29	11,67	5,74	11,15	11,23	3,84	1,06
300	1	8,7	9,44	12,02	9,83	12,58	5,91	10,72	11,46	5,12	1,03
50	2	5,67	7,54	10,85	6,11	9,72	7,03	11,77	10,67	4,71	0,77
100	2	5,15	7,85	10,63	5,94	9,98	5,44	10,66	10,21	3,41	0,83
200	2	6,6	8,57	11,23	7,76	11,5	5,43	10,49	10,8	3,62	0,98
300	2	8,44	9,35	11,76	9,61	12,56	5,73	10,37	11,15	5,05	1
50	5	4,64	6,92	9,73	4,91	9,05	5,73	9,5	9,13	3,91	0,59
100	5	4,62	7,51	10,14	5,33	9,73	4,87	9,47	9,24	3,11	0,73
200	5	6,24	8,37	11,01	7,52	11,42	5,06	9,98	10,3	3,51	0,94
300	5	8,17	9,19	11,65	9,48	12,54	5,45	10,13	10,95	4,97	0,97
50	10	4,3	6,72	9,41	4,59	8,76	5,35	8,78	8,59	3,74	0,52
100	10	4,5	7,34	9,96	5,2	9,61	4,76	9,08	8,96	3,07	0,7
200	10	6,21	8,26	10,91	7,48	11,35	4,98	9,85	10,2	3,46	0,91
300	10	8,1	9,1	11,54	9,47	12,5	5,42	10,04	10,97	4,92	0,95