



ESTIMATIVA DE MATRIZES DE ORIGEM-DESTINO A PARTIR DE
REGISTROS DE CHAMADAS DE TELEFONIA CELULAR

Júlio César Chaves

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

Orientador: Alexandre Gonçalves Evsukoff

Rio de Janeiro

Março de 2018

ESTIMATIVA DE MATRIZES DE ORIGEM-DESTINO A PARTIR DE
REGISTROS DE CHAMADAS DE TELEFONIA CELULAR

Júlio César Chaves

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 CIVIL.

Examinada por:

Prof. Alexandre Gonçalves Evsukoff, D.Sc.

Prof. Nelson Francisco Favilla Ebecken, D.Sc.

Prof^a. Beatriz de Souza Leite Pires de Lima, D.Sc.

Prof. Rômulo Dante Orrico Filho, D.Sc.

Prof. Artur Ziviani, Ph.D.

Prof. Moacyr Alvim Horta Barbosa da Silva, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2018

Chaves, Júlio César

Estimativa de matrizes de origem-destino a partir de registros de chamadas de telefonia celular/Júlio César Chaves. – Rio de Janeiro: UFRJ/COPPE, 2018.

XX, 177 p.: il.; 29, 7cm.

Orientador: Alexandre Gonçalves Evsukoff

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

Referências Bibliográficas: p. 144 – 153.

1. mobilidade. 2. urbana. 3. celular. I. Evsukoff, Alexandre Gonçalves. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Civil. III. Título.

*À minha querida mãe (in
memoriam), que desde cedo me
incentivou aos estudos na
esperança de vencer a pobreza. À
Santa Catarina de Alexandria,
protetora dos estudantes, que me
acompanha desde que aprendi a
rezar.*

Agradecimentos

Agradeço à paciência, amor e compreensão da minha esposa, sem a qual nada disso seria possível.

Agradeço ao convite, o incentivo e a paciência do meu orientador Alexandre, sem o qual esta pesquisa não teria iniciado.

Agradeço ao Professor Moacyr Silva da Escola de Matemática Aplicada pela abertura e coordenação do projeto de pesquisa de mobilidade urbana junto a rede de pesquisas da FGV, assim como ao Diretor Ricardo Simonsen pela aprovação e suporte ao projeto.

Agradeço à arguição da banca examinadora: Prof. Artur Ziviani, Prof. Rômulo Dante Orrico Filho, Prof^a. Betriz de Souza Leite Pires de Lima, Prof. Nelson Francisco Favilla Ebecken, e novamente aos Professores Moacyr e Alexandre. O pouco que possa ter ido além, foi porque estava “*sobre ombros de gigantes*”.

Agradeço à colega Ligia Alves da Cruz, assim como todos os colaboradores da Biblioteca Mario Henrique Simonsen (FGV-RJ), por manterem disponível tão diverso e excelente acervo, e estarem sempre prontos para auxiliar usuários e pesquisadores. Além de fonte de pesquisa, a biblioteca tornou-se uma parte da minha própria formação de pesquisador.

Agradeço ao colega Glauco Clemente Marques e à equipe de operações da FGV, pelo suporte operacional provido através da “cloud acadêmica”, que foi capaz de alavancar o desenvolvimento de uma pesquisa com tamanha volumetria de dados.

Agradeço ao incentivo e confiança de meus superiores na Diretoria de Operações da FGV, Mario Rocha, Ricardo Güth e José Ricardo Cavalcante.

Agradeço à Marta González (MIT), por compartilhar a experiência em modelagem de mobilidade na ocasião do meu exame de qualificação.

Agradeço à amiga Ona Kiser, pelas orações e pelas dicas de língua Inglesa.

Agradeço à amiga, colega de grupo vocal do Outeiro da Glória e Professora, Lucia Helena Salgado e Silva Pedra, pelo incentivo e compreensão, assim como a todos os amigos do grupo vocal e da Imperial Irmandade de Nossa Senhora da Glória do Outeiro.

De forma geral agradeço a paciência e peço desculpas a todos os amigos e parentes pela longa ausência de qualquer convívio social, sentida por muitos entes queridos.

Ao então secretário de transportes do Estado do Rio de Janeiro, Carlos Roberto Osório, que nos recebeu prontamente para uma apresentação do projeto de pesquisa e nos ofereceu os dados compilados pelo Plano Diretor de Transportes.

Agradeço à Rebecca Bittencourt da Secretaria Municipal de Transportes, pelo direcionamento na questão da largura das ruas do Rio de Janeiro, assim como a promoção de uma reunião na Secretaria, de onde vimos que o anseio de inovação no transporte público está de certa forma relacionado ao uso de dados de telefonia móvel. Espero que esta pesquisa seja útil à sociedade através dos colegas da área de transportes.

Em termos de mobilidade na Região Metropolitana do Rio de Janeiro, diversas pessoas colaboraram com suas *experiências pessoais de trânsito*, de onde surgiram algumas dicas muito profícuas. Em especial, agradeço aos colegas de trabalho da FGV: Edson Correia Marques, pelo *insight* sobre a obra do Arco Metropolitano, fazendo-me analisar a mobilidade através do Porto de Itaguaí; ao colega Otavio Freitas Teixeira, pela dica sobre Magé, que em feriados prolongados acaba servindo como passagem por diversas pessoas que tentam “fugir” dos engarrafamentos em direção à Região dos Lagos no Carnaval; aos colegas da FGV Projetos,

João Augusto Pereira Batista, Eduardo Santiago Rosseti, Baiena Feijolo Souto, Eduardo Leal de Oliveira e Gabriel Stumpf Duarte de Carvalho, pelas dicas e conceitos acerca de questões sócio-econômicas e da área de transporte; a todos os colegas da FGV que de alguma forma acompanharam-me nessa grande “maratona” que é uma tese.

Em termos de metodologia científica, tenho muito a agradecer à Dionísio Carmo-Neto, mesmo sem conhecê-lo, pois foi a partir de seu livro de *metodologia científica para principiantes*, que compreendi que a escrita e a leitura vão além de si mesmas. Compreendi que: *ler é como atravessar uma rua sob os olhos de outra pessoa; escrever também é atravessar uma rua, contudo sabendo que centenas de pessoas irão atravessar a mesma rua sob os seus olhos.*

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

ESTIMATIVA DE MATRIZES DE ORIGEM-DESTINO A PARTIR DE
REGISTROS DE CHAMADAS DE TELEFONIA CELULAR

Júlio César Chaves

Março/2018

Orientador: Alexandre Gonçalves Evsukoff

Programa: Engenharia Civil

Este trabalho apresenta um método de representação espacial por unidades geográficas geradas pela agregação de torres de telefonia celular. Esta representação geo-espacial facilita a manipulação (cruzamento) de múltiplas “visões” de dados. Uma base de dados de telefonia móvel foi coletada durante o ano de 2014 para estimar matrizes de origem-destino por dia, de onde foram extraídos padrões de mobilidade ordinários e extraordinários. A base do censo de 2010 foi usada para correções populacionais. A inferência da região de residência dos usuários possibilitou a criação de duas matrizes comuns à área de transporte: *Home-Based Others* (HBO) e *Non-Home Based* (NHB). A mudança nos hábitos de mobilidade da população foi detectada por conta de feriados prolongados. A base da área de transportes foi usada para validações quantitativas de viagens, e os resultados se mostraram compatíveis com as estimativas oficiais. O método com fator de correção simples apresentou uma tendência a subestimar o número de viagens, principalmente para valores pequenos. Ou seja, quanto maior a quantidade de dados disponíveis sobre a região de estudo, melhores se tornam as inferências.

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

ORIGIN-DESTINATION MATRIX ESTIMATION GIVEN BY CALL DETAIL
RECORDS

Júlio César Chaves

March/2018

Advisor: Alexandre Gonçalves Evsukoff

Department: Civil Engineering

This work presents a spatial method to partition geographic units given a set of cell phone towers. This geospatial method make easier to deal with multiple data layers. A cell phone call database from 2014 was used to estimate origin-destination matrices by day, covering ordinary and extraordinary mobility behavior. The Census data of 2010 was used to adjust population expansion factors. The user's residence inferring made possible the conception of two common matrices: *Home-Based Others* (HBO) and *Non-Home Based*(NHB). The population habits' changing was detected on long holidays. The transportation database was used to validate travels counting, and the results has showed to be compatible with the official estimations. The simple population expansion factor seems to overestimate the travels counting, mainly for regions with less data, i.e., better inferences come out when more cell phone data are available for the study area.

Sumário

Lista de Figuras	xii
Lista de Tabelas	xix
1 Introdução	1
1.1 Motivação	2
1.2 Contribuições	4
1.3 Estrutura da dissertação	5
2 Revisão Bibliográfica	7
2.1 Avaliação de transporte urbano	9
2.2 Dados de telefonia móvel	24
2.3 Análise e planejamento de eventos	36
2.4 Modelo de radiação	42
2.5 Tópicos conclusivos	44
3 O conjunto de dados	46
3.1 Os dados de CDR	50
3.2 Os dados do censo	56
3.3 Os dados do PDTU	59
4 Estimativas de matrizes de origem-destino	63
4.1 Padronização geográfica	65
4.2 Particionamento espacial	68

4.3	Domicílio presumido dos usuários de telefonia	69
4.4	Matrizes de origem-destino quotidianas	80
4.5	Matriz de origem-destino extraordinária	88
5	Resultados de matrizes O-D quotidianas	97
5.1	Discussão sobre o particionamento espacial	97
5.2	Validação dos fatores de expansão populacional	99
5.3	Validação da quantidade de viagens	101
5.4	Identificação de padrões de mobilidade	103
5.5	Detecção de mudança em padrões de mobilidade	116
6	Resultados de matrizes O-D extraordinárias	121
6.1	Validação da quantidade de pessoas em eventos	122
6.2	Visitantes nos blocos de carnaval em 2014	124
6.3	Visitantes no Reveillon	129
6.4	Visitantes das praias do Rio de Janeiro durante o verão	135
7	Conclusão e trabalhos futuros	141
	Referências Bibliográficas	144
A	Apêndice	154
A.1	Unidades geográficas de estudo	155
A.2	Instruções SQL ANSI	157
A.3	Guia de tecnologias usadas	172

Lista de Figuras

1.1	Fluxo metodológico geral de pesquisa.	5
2.1	A urbanização do mundo a partir de 2008 [1] e a projeção até 2050, segundo as Nações Unidas. Fonte: UN, weforum.org.	8
2.2	Exemplificação de cordão externo e cordão interno. O cordão externo é a linha imaginária (<i>em amarelo</i>) que delimita externamente a área de estudo. O cordão interno é a linha imaginária (<i>e.g. o rio em azul</i>) que possibilita a saída ou a entrada do cordão externo. . .	11
2.3	O modelo de quatro etapas, FSM. Fonte: Figura 2 de MCNALLY [2] traduzida ao português.	13
2.4	Comparativo percentual entre os modos de viagens. Fonte: Elaboração própria a partir de dados do PDTU de 2003 [3] e de 2013 [4].	23
2.5	Exemplo de área de cobertura das torres de telefonia com o uso de polígonos de voronoi. Fonte: Figura 1 de [5], foram feitas as traduções ao português.	29
2.6	Área de cobertura das torres de telefonia da região de estudo da presente tese, através de polígonos de voronoi. Fonte: Elaboração própria.	29
2.7	Um mesmo subdistrito é destacado em verde (Rocinha), mostrando a sobreposição entre a área de cobertura de torres de telefonia e um grid regular, imagens (a) e (b) respectivamente.	31

2.8	Mudança de localização através de uma rede de telefonia GSM, cada hexágono representa uma antena. Fonte CACERES <i>et al.</i> [6], traduzido ao português.	34
2.9	É visto um fluxo transiente entre B e C, quando na verdade o usuário partiu de A até D. Fonte TOOLE <i>et al.</i> [7], traduzido ao português.	35
2.10	Tipos de eventos segundo a perspectiva de turismo. Fonte: Elaboração própria, a partir de GETZ e PAGE [8].	38
2.11	Tipos de eventos segundo escala ou abrangência turística. Fonte: Elaboração própria, a partir de GETZ e PAGE [8].	38
2.12	População s considerada no modelo de radiação, excetuando-se os residentes de m e n . Fonte SIMINI <i>et al.</i> [9].	43
3.1	Área de estudo.	48
3.2	Distribuição de probabilidade de chamadas (a) e de vizinhos (b) cobrindo o ano de 2014 da base de CDR. Fonte: Elaboração própria.	51
3.3	Quantidade de chamadas realizadas por dia e por estação do ano. .	53
3.4	Visão geral da área de estudo com foco na distribuição espacial das antenas com seus respectivos polígonos de Voronoi. No detalhe a localização da Região Metropolitana do Rio de Janeiro.	53
3.5	Cidades da área de estudo. Fonte: Elaboração própria a partir do censo de 2010.	57
3.6	Cidades da área de estudo, sendo a cidade do Rio de Janeiro particionada em subdistritos. Fonte: Elaboração própria a partir do censo de 2010.	58
3.7	Renda domiciliar per capita das cidades da área 21, apenas a cidade do Rio de Janeiro está particionada em subdistritos. Fonte dos dados: censo de 2010.	58
3.8	Área incluída na pesquisa do PDTU de 2013 com destaque para os maiores corredores de transporte. Figura extraída da página 65 do PDTU 2013 [4].	60

3.9	Macrozonas de tráfego da RMRJ segundo o PDTU de 2013.	61
4.1	Quantidade de usuários ativos por hora (média de todos os dias). .	64
4.2	Quantidade média de chamadas processadas por torre de telefonia.	65
4.3	Recorte da área de estudo mostrando as antenas sobre alguns sub- distritos.	66
4.4	Quantidade de torres de telefonia por Km ² , por unidade geográfica do particionamento misto.	67
4.5	Quantidade média de chamadas por dia por região.	67
4.6	Particionamento espacial.	68
4.7	Visão do particionamento espacial, com ênfase nos locais em houve maior quantidade de registros.	69
4.8	Posicionamento metodológico do domicílio presumido.	70
4.9	Fluxo de passos para cálculo do domicílio presumido, desde o dado “bruto” de CDR até a quantidade de usuários residentes por unidade geográfica.	70
4.10	Quantidade de locais de permanência detectados sob os filtros es- tabelecidos, segundo quantidade de usuários (em escala logarítmica).	72
4.11	Diferença entre a quantidade total de detecções e quantidade de detecções no local inferido como residência.	74
4.12	Fatores de expansão populacional.	76
4.13	Dispersão de k_l fixo para as 55 unidades geográficas.	78
4.14	Dispersão de $\overline{k(OD)}$, média entre todos os dias úteis do ano, adap- tado para os $55 \times 55 = 3025$ pares OD	80
4.15	Fluxo metodológico das matrizes O-D quotidianas.	81
4.16	Fluxo de etapas para o cálculo geral de matrizes O-D quotidianas por dia, desde as chamadas realizadas por todos os usuários que tiveram a residência presumida até o resultado, que corresponde à matriz O-D $N \times N$	82

4.17 Fluxo de etapas para o cálculo geral de matrizes O-D quotidianas por dia, destaque para as condições que calculam as matrizes HBO e NHB, c.1 e c.2 respectivamente.	86
4.18 Posicionamento metodológico da matriz O-D extraordinária.	88
4.19 Fluxo de etapas para o cálculo da matriz O-D extraordinária.	89
4.20 Diferença entre a distância linear (linha preta) e a distância pelas vias de acesso (azul). Fonte: Google Maps.	92
4.21 Modelagem Entidade/Relacionamento (ER) para eventos.	93
5.1 Usuários de telefonia detectados em 2014 nas unidades geográficas de estudo <i>vs.</i> população declarada pelo censo 2010 para as mesmas unidades.	100
5.2 Comparação entre a quantidade de viagens inferida pelos fatores de expansão populacionais k-fixo e k-adaptativo <i>vs.</i> PDTU.	102
5.3 Comparação entre a média de viagens entre todos os pares <i>OD</i> em dias úteis, tanto via k-fixo quanto via k-adaptativo.	103
5.4 Quantidade de viagens zona sul x centro (k-fixo), visto através das matrizes NxN, NHB e HBO.	105
5.5 Padrão semanal Casa-Trabalho.	105
5.6 Média de chamadas por dia da semana de todos usuários com residência presumida.	106
5.7 Quantidade de viagens zona sul x centro, visto através da matrizes NxN obtida via k-fixo e via k-adaptativo.	107
5.8 Quantidade de viagens Centro x Zona Sul (k-fixo), visto através das matrizes NxN, NHB e HBO.	108
5.9 Quantidade de viagens Casa x Trabalho e Trabalho x Casa (k-fixo), equivalente a Centro x Zona Sul e Zona Sul x Centro, respectivamente, visto apenas pelas matrizes NxN.	109
5.10 Média de viagens NxN (k-fixo) por dia da semana entre subdistritos da zona sul e centro; e vice-versa.	109

5.11	Quantidade de viagens Zona Sul x Teresópolis e Teresópolis x Zona Sul (k-fixa), visto pela matriz $N \times N$	110
5.12	Média de viagens por dia da semana entre Zona Sul x Teresópolis e Teresópolis x Zona Sul (k-fixa), visto pela matriz $N \times N$	111
5.13	Quantidade de viagens Zona Sul x Teresópolis vista pela matriz $N \times N$, obtida via k-fixa e via k-adaptativo.	111
5.14	Principais eixos viários rumo à Teresópolis e Região dos Lagos. . .	112
5.15	Viagens Niterói x Maricá e Maricá x Niterói, visto pela matriz $N \times N$.	113
5.16	Viagens Niterói x Maricá, visto pelas matrizes $N \times N$ (laranja) e NHB (vermelho).	113
5.17	Quantidade de viagens Niterói x Maricá vista pela matriz $N \times N$, obtida via k-fixa e via k-adaptativo.	114
5.18	Quantidade de viagens Magé x Itaboraí e Itaboraí x Magé, visto pela matriz $N \times N$	115
5.19	Quantidade de viagens Magé x Cachoeiras de Macacu e Cachoeiras de Macacu x Magé, visto pela matriz $N \times N$	115
5.20	Quantidade de viagens Magé x Teresópolis e Teresópolis x Magé, visto pela matriz $N \times N$	115
5.21	Viagens (k-fixa) entre Duque de Caxias e Itaguaí (e vice-versa), média móvel de 15 dias anteriores. A barra em (a) indica a data de inauguração do Arco Metropolitano. A média de viagens/dia da semana, (b) e (c), comprovam maior aumento em dias úteis.	117
5.22	Quantidade de viagens Duque de Caxias x Itaguaí, visto pela matriz $N \times N$ obtida via k-fixa a via k-adaptativo.	118
5.23	Média móvel de 20 dias (10 dias anteriores e 10 dias posteriores) das viagens $N \times N$ Barra x Madureira (k-fixa).	119
5.24	Média móvel de 15 dias anteriores das viagens $N \times N$ Barra x Bangu (k-fixa).	119

5.25	Quantidade de viagens NxN Barra x Bangu obtidas via k-fixa e via k-adaptativo.	120
6.1	Mapa da cobertura do bloco Banda de Ipanema. O percurso publicado pelos organizadores está destacado através das ruas em amarelo, e a área de cobertura das torres de telefonia está destacada pelo polígono roxo.	124
6.2	Aparente falta de relação entre pessoas/ m^2 sob os dois tipos de área: útil e torres de telefonia.	127
6.3	Quantidade de visitantes aos principais blocos de carnaval de 2014 na cidade do Rio de Janeiro.	128
6.4	Probabilidade de um indivíduo ir um bloco de carnaval vs. a distância a percorrer a partir da própria residência.	129
6.5	Mapa das áreas consideradas em Copacabana. A Orla (incluindo o calçadão) está destacada em amarelo, e a área de cobertura das torres de telefonia está destacada pelo polígono azul.	132
6.6	Distribuição geográfica das pessoas que foram a algum evento de Reveillon em 2014 onde houve queima de fogos.	134
6.7	Volume de visitantes do Reveillon 2014, indicando a tendência de preferência por atrações próximas de casa.	135
6.8	Probabilidade versus distância percorrida de indivíduos de determinado local irem aos destinos de Reveillon. A escala é logarítmica.	136
6.9	Praias da orla do Rio de Janeiro (do Leme ao Pontal) delimitadas pela cobertura dos polígonos de voronoi das torres de telefonia correspondentes.	137
6.10	Média de visitas às praias da cidade do Rio de Janeiro durante todos os domingos do verão 2013/2014.	139
6.11	Média da probabilidade em função da distância percorrida das pessoas irem às praias da cidade do Rio de Janeiro durante todos os domingos do verão 2013/2014.	140

A.1	Figura gerada com o software QGIS.	174
A.2	Figura gerada com o software Tableau. Obtida a partir do site de exemplos públicos da própria Tableau.	176
A.3	Mapa gerado pelo mapbox.	176

Lista de Tabelas

2.1	Exemplo de organização de dados sócio-econômicos por zona de tráfego. Fonte: tabela 2 de [2].	12
2.2	Exemplo de organização de dados demográficos dos domicílios. Quantidade de domicílios por zona e por quantidade de carros, assim como pela faixa de renda. Fonte: recorte da tabela 3 de [2].	12
2.3	Exemplo de características de uma rede de transportes. Fonte: recorte da Tabela 1 de [2].	12
2.4	Siglas internacionais para tipificar intervalos de tempo padrão e motivos de viagens. Fonte: recorte da Tabela 4.4.1 na página 168 em [4].	14
3.1	Percentual de pessoas com telefone móvel celular para uso pessoal, na população de 10 anos ou mais de idade (%) por grandes regiões. Fonte: PNAD IBGE – Pesquisa Nacional por Amostra de Domicílios.	47
3.2	Cidades da área de estudo.	49
3.3	Dados de um registro de chamada telefônica (CDR)	50
3.4	Contadores obtidos após primeiro pré-processamento, contemplando todo o ano de 2014.	52
3.5	Estatísticas das torres de telefonia nas cidades contempladas na área de estudo.	55

6.1	Visitantes nos blocos de carnaval do Rio de Janeiro em 2014 segundo estimativas da metodologia de CDR.	127
6.2	Visitantes nos locais típicos de Reveillon na RMRJ em 2014, segundo estimativas da metodologia de CDR.	132
6.3	Visitantes nos locais típicos de Reveillon na RMRJ em 2015, segundo estimativas da metodologia de CDR.	133
6.4	Principais praias do Rio de Janeiro com suas respectivas áreas de orla, cobertura total das torres de telefonia e cobertura média por torre.	137
6.5	Visitantes às praias do Rio de Janeiro no verão 2013/2014, segundo estimativas da metodologia de CDR.	138
A.1	Unidades do particionamento misto adotadas para a área de estudo, relacionadas às macrozonas de tráfego da RMRJ.	155
A.1	Unidades do particionamento misto adotadas para a área de estudo, relacionadas às macrozonas de tráfego da RMRJ.	156
A.1	Unidades do particionamento misto adotadas para a área de estudo, relacionadas às macrozonas de tráfego da RMRJ.	157

Capítulo 1

Introdução

“Uma obra de arte é boa quando nasce por necessidade.”

— Rainer Maria Rilke, Cartas a um jovem poeta.

Conteúdo

1.1	Motivação	2
1.2	Contribuições	4
1.3	Estrutura da dissertação	5

Os desafios enfrentados pelos engenheiros de transporte podem estar ligados aos desafios das grandes cidades em lidar com a forte urbanização. Em 1965, DAVIS [10] previu que mais da metade da população mundial estaria vivendo em grandes centros até 1990, isso concretizou-se em 2008¹. No Brasil, o censo do ano 2000 revelou que 84% da população já era urbana. Como possíveis reflexos, os moradores das grandes cidades habituaram-se aos longos engarrafamentos. Estima-se que a “conta” dos engarrafamentos nos E.U.A chegou aos US\$ 160 bilhões em 2014 [11]. Cerca de um milhão de pessoas leva mais de duas horas até

¹Segundo UNITED NATIONS DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS/POPULATION DIVISION [1] a população urbana mundial igualou-se à rural em 2008.

o trabalho no Brasil², imaginando a volta para casa, contam-se duas horas adicionais, ou seja, no mínimo quatro horas gastas no trânsito por dia de trabalho.

Os inconvenientes das metrópoles parecem agravar-se na ocasião de eventos, tais como Carnaval e Reveillon, quando longos engarrafamentos podem ser observados na cidade do Rio de Janeiro, por exemplo. Apesar de existir um plano de transporte urbano que contemple essa região, o plano não abrange a questão de mobilidade por conta de eventos. Mesmo que abrangesse, a periodicidade de atualização do plano, 10 anos, pode não ser suficiente para acompanhar as mudanças na mobilidade da população. Essa é uma das motivações em usar dados de telefonia móvel para inferir a geração de viagens, conforme a seguir.

1.1 Motivação

A principal motivação desta pesquisa é oferecer uma metodologia para inferência de mobilidade da população com baixo custo, a partir da formatação e uso dos dados de telefonia móvel. O uso de telefone celular vem aumentando nos últimos anos e os dados de registros de chamadas, que normalmente são jogados fora, podem ser úteis e até trazer benefícios à sociedade. O número de usuários de celular cresceu de 12% da população mundial no ano 2000 para 96% em 2014; são 6,8 bilhões de assinantes [12]. O caso não é diferente no Brasil, onde 36,6% da população possuía celular em 2005 e 78,3% em 2015³.

A popularização dos telefones celulares já vêm estimulando o uso de serviços baseados em geo-localização e em padrões de comportamento. A gama de serviços é vasta, é possível saber quantos passos se andou durante o dia (Google Fit), receber alertas de trânsito em tempo de tomar alguma medida ao volante (Waze), ver onde está ocorrendo troca de tiros no Rio de Janeiro (fogo cruzado), apenas para citar alguns exemplos.

²Censo de 2010.

³Fonte: IBGE, Diretoria de pesquisas, coordenação de trabalho e rendimento, pesquisa nacional por amostra de domicílios.

O avanço no uso e na popularização da telefonia móvel vem acompanhado de uma grande demanda por armazenamento de dados, já que os aparelhos começam a tornar-se uma espécie de “sensor” pessoal. O Facebook, por exemplo, armazena 35TB por hora⁴, algo equivalente a mais de 60mil CDs em apenas uma hora⁵. Esse novo tipo de demanda de armazenamento possui alguns termos mencionados, tais como “Big Data” e “Data Lake”⁶. Sob um enfoque voltado para soluções inteligentes que trazem benefícios à sociedade, o termo “smart cities” vem sendo usado.

O termo que define os dados de celular usados nesta pesquisa é o CDR, que advém do inglês *Call Detail Records*, e refere-se ao registro telefônico de uma chamada de um indivíduo num único instante de tempo para um único aparelho. O registro indica principalmente a data, hora em que a chamada começou, a duração, o número de origem, o número de destino, o número da antena que atendeu a chamada, e o tipo de chamada: local, interurbana, internacional, etc. A motivação deste estudo é explorar os dados de CDR e realizar agregações com bases auxiliares para extrair e organizar diversas matrizes de origem-destino, dando suporte à área de transportes e fomentando a construção de “smart cities”.

Os aglomerados urbanos estão sujeitos a engarrafamentos, poluição e um intenso grau de contato humano e social [10], de forma que:

“As cidades precisam ser compreendidas como organismos vivos e complexos, repletos de dados não-examinados, contudo obviamente interconectados, certamente compreensíveis e relacionados”.(JACOBS [13]).

Tradicionalmente, a área de transportes realiza estudos de mobilidade através de entrevistas, contando com a participação voluntária e direta da população, esta pesquisa prevê a participação involuntária e indireta, através dos dados de CDR.

⁴Visto em 2014.

⁵Segundo a conferência Vertica Big Data de 2014, ver <http://vertica.tips/2014/08/27/how-vertica-met-facebooks-35tb-per-hour-ingest-sla/>

⁶Grande repositório de dados.

1.2 Contribuições

O objetivo deste estudo é apresentar uma metodologia para estimativa de matrizes de origem-destino a partir de dados de CDR. Apesar da pesquisa ter sido desenvolvida considerando a Região Metropolitana do Rio de Janeiro (RMRJ) como área de estudo, o objetivo é que a metodologia seja aplicável a qualquer região. Usualmente, as matrizes de origem-destino (matrizes O-D) constam como parte integrante dos planos de mobilidade urbana elaborados pelo setor de transportes.

Esta pesquisa se propõe a usar os dados de CDR e censo demográfico (ver Fig.1.1) para:

- i)* Estabelecer um particionamento geográfico aderente às torres de telefonia e ao aspecto sócio-econômico da população;
- ii)* Estabelecer critérios de extração, a partir dos dados, do que vem a ser considerado uma "viagem" ou um "ponto de permanência";
- iii)* Inferir a região de residência dos usuários de telefonia a partir dos pontos de permanência detectados;
- iv)* Estabelecer fatores de expansão populacionais por região e por atividade de usuários de telefonia;
- v)* Apresentar a quantidade de viagens sob o formato de matrizes O-D quotidianas e extraordinárias, uma focada em comportamentos regulares durante o ano, subdividindo-se em HBO e NHB, e outra focada em grandes eventos, tais como Carnaval e Reveillon.

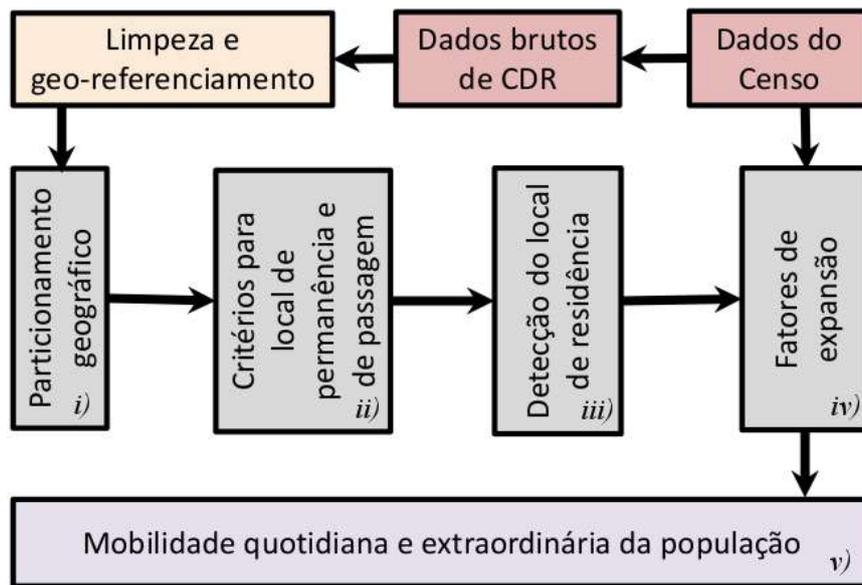


Figura 1.1: Fluxo metodológico geral de pesquisa.

Considerando as pesquisas abordadas na revisão bibliográfica (Capítulo 2), e os temas abordados na Figura 1.1, as principais contribuições desta tese consistem nos itens a seguir:

- a) Particionamento geográfico ao mesmo tempo aderente ao censo e aos dados de CDR;
- b) Fator de expansão populacional "adaptativo", considerando a atividade diária dos usuários de telefonia;
- c) Matrizes O-D quotidianas obtidas a partir de dados de CDR e subdivididas em HBO e NHB;
- d) Matrizes O-D voltadas para eventos extraordinários na RMRJ.

1.3 Estrutura da dissertação

A apresentação do estudo será subdividido da forma a seguir:

- Um capítulo de revisão bibliográfica, abordando conceitos e relatórios próprios da área de transportes, assim como iniciativas de uso de dados de

telefonia móvel nessa e outras áreas afins. Referências da área de turismo serão usadas para posicionar os conceitos sobre eventos, assim como referências de estabelecimento de modelos de transporte.

- Um capítulo de materiais, apresentando a área de estudo, os dados de telefonia móvel, os dados do censo, e os dados do Plano Diretor de Transporte, estimulando a discussão sobre um particionamento geográfico padronizado.
- Um capítulo de métodos, estabelecendo um modelo geográfico de pesquisa, de onde será possível, através de algoritmos, presumir o local de residência dos usuários de telefonia, inferir matrizes de origem-destino ordinárias e extraordinárias.
- Dois capítulos de resultados, apresentando as matrizes de origem-destino quotidianas, juntamente com uma forma de validação a partir de dados oficiais da Secretaria de Transporte.
- No segundo capítulo de resultados, serão apresentadas as matrizes de origem-destino extraordinárias, sob o contexto de alguns dos eventos mais comuns no Rio de Janeiro: Reveillon, Carnaval e visitas às praias durante o verão. Será apresentado um direcionamento para validação dessas matrizes, envolvendo uma contagem de pessoas por m^2 .
- Um capítulo de conclusão e trabalhos futuros, resumindo os resultados atingidos, a limitação da metodologia e uma proposta de continuidade.

Capítulo 2

Revisão Bibliográfica

*“A maior parte dos acontecimentos
é inexprimível e ocorre num
espaço em que nenhuma palavra
jamais pisou.”*

— Rainer Maria Rilke, Cartas a um
jovem poeta.

Conteúdo

2.1	Avaliação de transporte urbano	9
2.1.1	Modelo de quatro etapas - FSM	10
2.1.2	Planos de transporte urbano	18
2.2	Dados de telefonia móvel	24
2.2.1	Matriz O-D	32
2.3	Análise e planejamento de eventos	36
2.3.1	Uso de dados de CDR para análise de eventos	39
2.4	Modelo de radiação	42
2.5	Tópicos conclusivos	44

O estudo de mobilidade humana tem registros desde o século XIX, quando um cientista social usou as leis da física para comparar a tendência de as pessoas viverem próximas umas das outras com a força da gravidade. A partir de um estudo sobre ferrovias e migração, Henry Charles Carey observou que o ser humano tem a “tendência de ficar próximo de seus amigos”, a partir do original

em inglês, “gravitate to his fellow man” [14]. A força de atração é diretamente proporcional à quantidade de amigos e inversa à distância dos mesmos¹.

A partir de 2008, a população urbana mundial igualou-se à rural [1]. A figura 2.1 mostra essa marca e uma projeção até 2050², de onde se vê claramente uma estagnação da população rural e um aumento constante da população urbana.

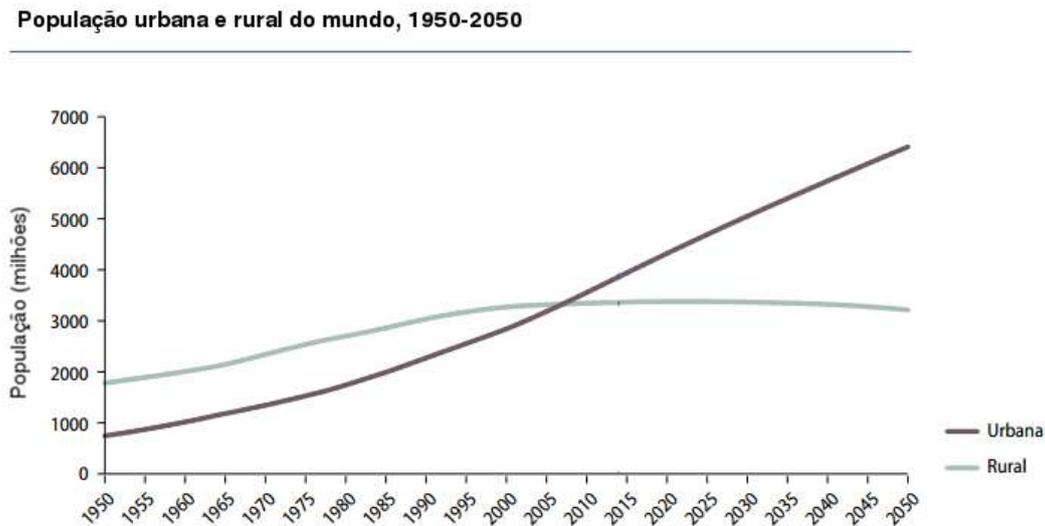


Figura 2.1: A urbanização do mundo a partir de 2008 [1] e a projeção até 2050, segundo as Nações Unidas. Fonte: UN, weforum.org.

O intenso grau de contato humano trouxe ao mundo urbano os inconvenientes das grandes cidades. Estima-se que a “conta” dos engarrafamentos nos E.U.A somaram US\$ 160 bilhões em 2014 [11]. Para realização de estudos, a área de planejamento de transporte urbano tem coletado dados através de questionários sobre padrões de viagens dos usuários, pesquisas domiciliares, contagens volumétricas de tráfego, informações de uso do solo e rede de transporte. A coleta de dados para o planejamento de transporte é custosa e demanda um longo período de planejamento e execução, e também possui um tamanho amostral limitado, tendo em vista os custos envolvidos [16].

¹Para uma breve descrição histórica, ver o capítulo introdutório de SEN e SMITH [15].

²<https://www.weforum.org/agenda/2016/10/the-world-s-10-largest-cities-by-2030> visto em 17 de agosto de 2017.

Uma pesquisa no catálogo de dissertações e teses da CAPES³ com os termos telefonia AND mobilidade⁴, mostra 71 trabalhos, após filtro na busca para trazer apenas as teses de doutorado, restaram apenas 9. Dessas 9, apenas 2 estão ligadas à engenharia, sendo as demais às ciências sociais, humanas, da terra e letras. Dessas 2, nenhuma delas trata a questão de mobilidade da população no sentido de sistemas de transporte. Se aplicar o filtro para dissertações de mestrado, são encontrados trabalhos relacionados a caracterização de mobilidade da população a partir de dados de celular: XAVIER [17], que usa dados de telefonia móvel para analisar a mobilidade urbana junto à carga de trabalho das antenas; e JOSÉ [18], que analisa a mobilidade da população por conta de eventos, uma comparação com alguns resultados de [18] será mostrada no capítulo de resultados. Assim sendo, trata-se de um tema relativamente novo no país.

A seguir serão apresentados os conceitos básicos usados pela área de transportes, assim como um resumo dos relatórios de transporte de alguns dos maiores centros urbanos do mundo e do país. Também serão apresentados alguns conceitos essenciais para compreender a representatividade dos dados de telefonia móvel, assim como a contextualização necessária para usá-los como insumo de pesquisa para transporte. Também serão vistos alguns conceitos básicos de eventos segundo a área de turismo, e como algumas pesquisas já identificaram oportunidades em usar os dados de celulares nessa área.

2.1 Avaliação de transporte urbano

Historicamente a modelagem de transportes se desenvolveu nos E.U.A. nos anos 50, aplicada às cidades de Detroit e Chicago [19]. A partir de 1970, se iniciaram as preocupações ambientais e a necessidade de um planejamento multimodal, o que vem tornando o estudo de mobilidade complexo, dadas as diversas interconexões e opções de transporte entre uma origem e um destino. Recentemente, diversos

³Ver [http://catalogodeteses.capes.gov.br/catalogo-teses/#!/.](http://catalogodeteses.capes.gov.br/catalogo-teses/#!/)

⁴Pesquisa feita em 5/12/2017.

estudos vêm adotando conceitos de redes complexas [20] para compreender a mobilidade da população [21]. Nesses estudos, geralmente os vértices da rede são vistos como zonas de tráfego e as arestas como vias de acesso entre uma zona e outra, sendo os modos de transporte disponíveis, as múltiplas camadas da rede [22], por exemplo. Embora a modelagem de transporte seja um tema amplo [23], geralmente o *modelo-base* adotado é o FSM (*four-step model* ou modelo de quatro etapas), sendo a matriz de origem-destino um dos produtos dessa metodologia.

As matrizes de origem-destino são instrumentos úteis à área de planejamento de transportes. Basicamente, elas servem à finalidade de organizar e compreender o volume de viagens entre todas as regiões de uma determinada área de estudo. Dentre as etapas do FSM, uma delas é responsável pela “geração de viagens” e outra pela “distribuição de viagens”, justamente sob a organização de matrizes de origem-destino. O posicionamento deste estudo reside principalmente na “geração de viagens”, usando os dados de telefonia móvel como indicadores de viagens.

A seguir será apresentada uma breve descrição do modelo FSM, e logo após, alguns planos elaborados pelas áreas de transporte de grandes centros urbanos.

2.1.1 Modelo de quatro etapas - FSM

O FSM serve basicamente à finalidade de organizar e gerir as demandas de transporte, sob os seguintes *pré-requisitos*: *a)* a delimitação da área de estudo através do *cordão externo* (ver *Figura 2.2*), *b)* dados sobre a população residente na área de estudo, e *c)* dados sobre todo o sistema viário disponível dentro do cordão externo. A seguir uma breve descrição desses itens.

Tabela 2.1: Exemplo de organização de dados sócio-econômicos por zona de tráfego. Fonte: tabela 2 de [2].

Zona interna	Total de residências por zona	Total de empregados por zona			
		Varejo	Serviços	Outros	Total
1	1400	800	400	800	2000
2	1200	800	400	400	1600
3	800	200	400	200	800
4	600	200	200	0	400
Total	4000	2000	1400	1400	4800

Tabela 2.2: Exemplo de organização de dados demográficos dos domicílios. Quantidade de domicílios por zona e por quantidade de carros, assim como pela faixa de renda. Fonte: recorte da tabela 3 de [2].

	Zona 1			Zona 2		
	Baixa renda	Média renda	Alta renda	Baixa renda	Média renda	Alta renda
sem carro	40	80	80	20	40	40
1 carro	120	320	360	80	260	160
2 carros	40	200	160	100	300	200

mada de *sistema de transporte*, pode ser vista na Tabela 2.3. A disposição dos pré-requisitos e as etapas do FSM podem ser vistas na Figura 2.3, a partir de MCNALLY [2]. Observe que o sistema de atividades busca compreender tudo o que for relativo às pessoas, às atividades humanas, ao passo que o sistema de transporte busca caracterizar tudo o que for relativo à infraestrutura viária.

Tabela 2.3: Exemplo de características de uma rede de transportes. Fonte: recorte da Tabela 1 de [2].

Tipo de conexão (todas de 1 via)	Velocidade (km/h)	Número de linhas	Capacidade por linha	Capacidade (veículos/hora)
1 freeway	120	2	200	400
2 arteria primária	120	2	100	200
3 rua coletora	60	1	100	100
4 conector central	30	9	100	900

Geralmente o sistema viário é uma rede complexa. O FSM foi desenvolvido para tratar essa complexidade através de um processo sequencial executado em quatro etapas (*ver Figura 2.3*). Na geração de viagens, (1) a frequência de viagens é obtida através da propensão em viajar. Os destinos, as produções e as atrações de viagens, são estimadas em tabelas de viagens separadamente. Na distribuição

de viagens, (2) os polos atratores, os polos produtores e os destinos das viagens são distribuídos de forma à refletir a impedância (custo e/ou tempo) da viagem. Na escolha modal, (3) as viagens são calculadas para que reflitam de forma proporcional os modos disponíveis. Na escolha da rota, (4) as tabelas modais de viagens são associadas à rede de transporte específica de cada modo.

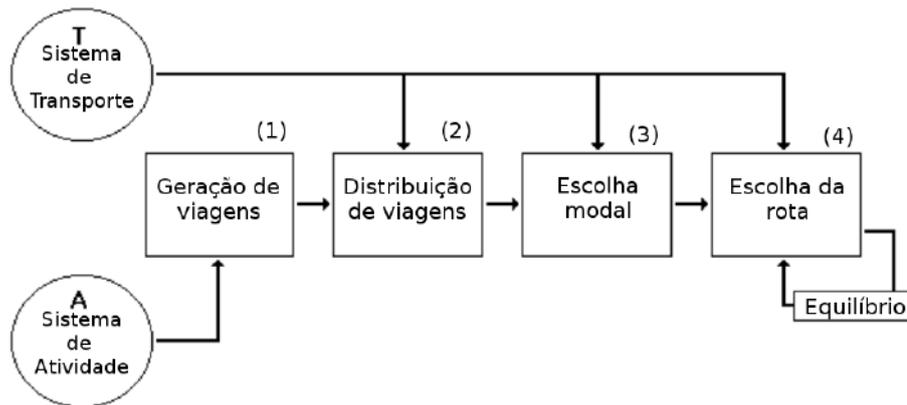


Figura 2.3: O modelo de quatro etapas, FSM. Fonte: Figura 2 de MCNALLY [2] traduzida ao português.

O FSM [2], na forma básica, prevê a alocação de viagens de forma equilibrada entre oferta e demanda apenas na etapa da escolha da rota. Ou seja, a demanda total é geralmente considerada fixa nas três primeiras etapas e ajustável na última, dependendo da capacidade de cada rota (conforme exemplo da Tabela 2.3).

A hora do dia geralmente é introduzida depois da etapa de distribuição de viagens, onde as tabelas de *produção-atração* de viagens são calculadas para refletir a distribuição de viagens ao longo das horas do dia, e posteriores agregações temporais para refletir intervalos de pico, etc. Geralmente são padronizadas siglas para definir esses intervalos de tempo, e até mesmo os motivos das viagens. A Tabela 2.4 mostra algumas dessas siglas [4].

A seguir uma breve descrição de cada etapa do FSM.

Tabela 2.4: Siglas internacionais para tipificar intervalos de tempo padrão e motivos de viagens. Fonte: recorte da Tabela 4.4.1 na página 168 em [4].

Sigla	Significado
HW	Viagens individuais em veículos. Automóveis (motoristas e passageiros) e táxis (passageiros).
PT	Viagens em transporte público. Todos os modos.
AM	Pico da manhã.
IP	Período entre picos.
PM	Pico da tarde.
HBW	Viagens de casa com o objetivo trabalho.
HBS	Viagens de casa com o objetivo estudo.
HBO	Viagens de casa com qualquer objetivo.
HBA	Viagens de casa com todos objetivos (all).
NHB	Viagens de qualquer lugar, exceto casa, com qualquer objetivo.
Other	Viagens de qualquer lugar, com qualquer objetivo.

Geração de viagens

Esta é a maior etapa do FSM. Nesta etapa, as viagens são divididas entre os polos produtores e os atratores, observando-se os motivos. As Características do Sistema de Atividades (CSA) dos indivíduos são formatadas em função de viagens de polos produtores $f_P^p(CSA)$, e viagens para polos atratores $f_A^p(CSA)$, para cada motivo p , da seguinte forma:

$$\text{Polo Produtor } P_i^p = f_P^p(CSA) \quad (2.1)$$

e

$$\text{Polo Atrator } A_j^p = f_A^p(CSA) \quad (2.2)$$

onde: P_i^p é o total de viagens produzidas sob o motivo p a partir da zona de tráfego i , e A_j^p é o total de viagens atraídas sob o motivo p para a zona de tráfego j .

Os três tipos mais básicos de viagem são: casa-trabalho (HBW *home-based work*), casa-escola (HBS *home-based school*), casa-outros (HBO *home-based others*) e outros menos casa NHB (*non-home based*).

As viagens são modeladas sob a ótica de produção e atração. Ao mesmo tempo que uma viagem é produzida por um local, ela é atraída por outro. Po-

de-se observar alguns exemplos: *a)* uma região densamente povoada, em que as muitas pessoas precisem se locomover, seja por motivos de trabalho, estudo ou lazer, pode ser uma região produtora de viagens; *b)* uma universidade, que recebe milhares de alunos, professores e colaboradores por dia, pode ser uma região atratora de viagens. Ou seja, geralmente a região de residência é considerada um polo produtor de viagem, enquanto os demais destinos são considerados polos atratores. Importante observar que, mesmo que a pessoa esteja *voltando para casa*, o polo produtor continua sendo a região de residência, pois a viagem não teria sido *produzida*, e conseqüentemente a pessoa não precisaria *voltar para casa*, se não fosse pela atração no destino.

Ao final da geração de viagens, deve-se saber quantas viagens iniciaram e quantas finalizaram em cada zona de tráfego. Métodos alternativos vêm sendo usados para inferir a produção de viagens através de dados de celular, referências pertinentes a esse tema serão vistas à frente.

Distribuição de viagens

Esta é a segunda maior etapa do FSM, o objetivo dela é estabelecer o relacionamento matemático entre os pares *origens-destinos* das viagens. Para cada motivo p de viagem há uma tabela de viagens T_{ij} em função das características do sistema de atividades, e atributos da rede de transporte [2]. A área de pesquisas de transportes tem considerado formas alternativas para inferir origens e destinos, seja através de dados de contagem de tráfego [24], de dados de cartões de embarque (tipo bilhete único) [25], e a partir de dados de dispositivos móveis [26]. A forma tradicional da etapa de distribuição de viagens é:

$$T_{ij} = f_{TD}(P_i, A_j, F_{ij}) \quad (2.3)$$

onde: T_{ij} é a quantidade de viagens entre i e j ,
 F_{ij} representa o indicador de impedância (ou fricção) da viagem (tempo ou custo) entre as zonas i e j ,

P_i representa a quantidade de viagens produzidas na zona i ,

A_j representa a quantidade de viagens atraídas até a zona j (por simplificação o índice p , de tipo de viagem, foi omitido), e

f_{TD} é a função de distribuição de viagens (*trip distribution*), que geralmente segue o modelo gravitacional, conforme a seguir.

O modelo gravitacional é um procedimento matemático baseado na teoria gravitacional da física Newtoniana. A teoria de Newton prevê que a atração entre os corpos é diretamente proporcional à massa e inversamente proporcional à distância entre os mesmos. Esses conceitos obtiveram aderência à área de transportes, substituindo-se os corpos pela quantidade de viagens produzidas ou atraídas em áreas geográficas ou zonas de tráfego [27], já que observou-se a “tendência de o ser humano ficar próximo dos amigos”.

Matematicamente, o modelo gravitacional aplicado à distribuição de viagens, é definido por:

$$T_{ij} = P_i \left(\frac{A_j F_{ij} K_{ij}}{\sum_{k=1}^{zonas} A_k F_{ik} K_{ik}} \right) \quad (2.4)$$

onde K_{ij} é um fator opcional de ajuste para viagens entre as zonas i e j , essa variável geralmente é usada como um fator de expansão.

A elaboração do fator de impedância F_{ij} pode ser inferida manualmente para cada par ij , ou pode contar com o auxílio de uma função *gamma* para produzir uma curva suave e contínua [27]:

$$F_{ij} = a \times t_{ij}^b \times e^{c \times t_{ij}} \quad (2.5)$$

onde F_{ij} representa o indicador de impedância entre i e j ,

a , b e c são os coeficientes do modelo, sendo a um fator de expansão que pode variar sem mudar a distribuição,

t_{ij} é o tempo de viagem entre i e j , e

e a base logarítmica natural.

O resultado desta etapa consiste em transformar as viagens de um formato P-A (*produção-atração*) para um formato O-D (*origem-destino*), já que a próxima etapa recebe como entrada esse último formato.

Escolha modal

Apesar de tratar-se da etapa mais complexa do FSM [27], esta etapa pode ser facilitada através do uso de dados de bilhetes de ônibus, barcas, trens, metrô, pedágios, contagem de veículos, etc. Esses dados podem servir como ponto de partida para calibração de um modelo usado para essa finalidade [28].

Na análise da escolha modal, o total de viagens entre as zonas é sub-dividido em viagens realizadas sob cada modo disponível entre cada par O-D. A maioria dos modelos de escolha de modo são baseados no modelo *Logit*, usualmente adotado pela área de transportes para inferir a escolha de cada indivíduo em fazer ou não uma viagem.

Em alguns casos a etapa de escolha modal pode ser realizada antes da etapa de distribuição de viagens.

Escolha da rota

A associação de viagens à rede de transportes é a etapa final da modelagem e serve de base para validação, tanto na capacidade de replicar as viagens observadas como em avaliar melhorias futuras no sistema de transporte.

As viagens, segundo os modos, geralmente são distribuídas sob a premissa de *equilíbrio*, em que todas as rotas possuem a mesma impedância em momentos fora de pico. Os algoritmos usados para essa finalidade estão voltados a buscar o equilíbrio do tráfego através da descoberta de caminhos mínimos, o algoritmo Frank-Wolfe é um exemplo⁵.

⁵Ver https://en.wikibooks.org/wiki/Fundamentals_of_Transportation/Route_Choice visto em 28 de agosto de 2017.

Enfim, a estrutura básica do modelo de quatro etapas é a modelagem matemática do comportamento de viagens das pessoas, de forma a estimar e avaliar ações de planejamento.

Algumas das áreas de transporte responsáveis pelos maiores centros urbanos do mundo, têm institucionalizado a estrutura do FSM como parte de um extenso relatório técnico, direcionado à avaliação e à compreensão da mobilidade da população em seus aglomerados urbanos. Esses estudos possuem diferentes nomes dependendo do local a que foram direcionados, contudo os objetivos são basicamente os mesmos, variando apenas no grau de maturidade em que se encontram, enquanto alguns buscam avaliar se a demanda *versus* a oferta de opções de transporte estão compatíveis, outros mais avançados planejam redução de poluentes e criação de áreas verdes.

Os planos de transporte podem ser encontrados sob diversos nomes, no caso da RMRJ a referência é o Plano Diretor de Transporte Urbano (PDTU). Esses planos geralmente são extensos, abrangentes, e custosos. Serão apresentadas a seguir algumas dessas referências técnicas, cobrindo alguns dos maiores centros urbanos do mundo e do Brasil.

2.1.2 Planos de transporte urbano

A cidade de Tóquio, capital do Japão, é o maior centro urbano do mundo, e segundo projeções das Nações Unidas⁶, continuará sendo em 2030, com cerca de 37,2 milhões de habitantes. O escritório de desenvolvimento urbano de Tóquio⁷ apresenta uma visão básica de políticas que guiam as estratégias de implementação de melhorias que mantêm a cidade competitiva em relação aos padrões internacionais. Essas políticas, apenas para citar algumas, incluem: (1) restaurar espaços urbanos próximos de áreas verdes e fontes de água; (2) criar, transmitir e ensinar aos mais novos a cultura urbana única de Tóquio; (3) criar uma cidade

⁶Ver [weforum.org https://www.weforum.org/agenda/2016/10/the-world-s-10-largest-cities-by-2030](https://www.weforum.org/agenda/2016/10/the-world-s-10-largest-cities-by-2030), visto em 17 de agosto de 2017 e [29].

⁷*The bureau of Urban Development.*

onde os residentes possam viver confortavelmente, com segurança e com paz de espírito; e (4) obter e coordenar a participação de diversas entidades⁸ em prol do objetivo em comum [30].

As estratégias básicas adotadas pelo escritório de desenvolvimento urbano de Tóquio, usadas para a implementação das políticas citadas acima são: (a) melhorar a infraestrutura de transportes regionais; (b) estabelecer centros que aumentem a vitalidade econômica local; (c) fazer a transição para uma cidade com baixas emissões de carbono; (d) formar “corredores verdes” dentro da cidade; (e) criar espaços urbanos atrativos; e (f) criar uma cidade altamente segura e preparada para desastres.

A cidade de Delhi, capital da Índia, é o segundo maior centro urbano do mundo [29] e, segundo as projeções das Nações Unidas, continuará sendo o segundo em 2030, com cerca de 36 milhões de habitantes. O Ministério de Desenvolvimento Urbano da Índia apresentou em 2009 um relatório chamado “Functional Plan on Transport for National Capital Region” [31] cobrindo 115 zonas de tráfego, onde foi apresentado um planejamento desse centro urbano para o ano 2021. Alguns dos itens abordados são: (1) volume sobre capacidade das vias de acesso, assim como a respectiva velocidade média; (2) condições de pavimentação das vias, classificadas em boa, regular e ruim; (3) avaliação de toda a conectividade da rede viária, possibilitando a identificação da necessidade de corredores dedicados; e (4) avaliação dos níveis de ruído e poluição gerado pelas vias. O sistema de transporte aéreo também foi avaliado.

Ao final do plano funcional de transportes da cidade de Delhi, foram apresentadas diversas propostas de melhoria do sistema viário, tais como: (a) a criação de corredores dedicados para transporte de cargas e para atender áreas industriais; (b) a criação de diversas vias expressas buscando aliviar vias mais congestionadas; (c) a criação de mais um aeroporto; e (d) a criação de integrações multimodais. Também foi apresentado um plano de prevenção, redução e mitigação

⁸Residentes, empresas e organizações sem fins lucrativos.

de acidentes. A prevenção foi obtida principalmente através do planejamento das interconexões, a redução foi obtida pela conservação do sistema viário e a mitigação foi obtida através da monitoração (patrulhamento) das vias.

A elaboração de planos de transporte também é uma prática em alguns centros urbanos do Brasil. A lei federal 10.257 de 10 de julho de 2001 determina que cidades com mais de 500 mil habitantes elaborem um plano de transportes e trânsito. A lei federal 12.587 de 3 de janeiro de 2012 determina que municípios acima de 20 mil habitantes devem elaborar um plano de mobilidade urbana. Esses planos são encontrados sob diferentes nomes, como por exemplo: (1) Plano Diretor de Transporte Urbano (PDTU) para o Rio de Janeiro e Brasília; (2) Plano de Mobilidade Urbana e Transporte Integrado, para Curitiba; (3) Plano Diretor de Mobilidade Urbana (PlanMob), para São Paulo, Porto Alegre. Uma breve descrição de alguns desses relatórios técnicos será mostrada a seguir.

A região metropolitana de São Paulo é a maior do Brasil e a quinta maior do mundo, com uma população de mais de 21 milhões de pessoas⁹. O PlanMob 2015 de São Paulo mostra diversos indicadores e apresenta um planejamento para os próximos 15 anos. O relatório levou 3 anos para ser elaborado, entre 2013 e 2015, e aborda diversas estatísticas relativas aos (1) meios de transporte, (2) à emissão de gases poluentes, e (3) às mortes decorrentes de acidentes no trânsito. Essas estatísticas serão usadas para embasar as políticas públicas de transporte [32]. Serão mostrados a seguir alguns indicadores e ações correspondentes.

Segundo o PlanMob 2015 de São Paulo [32], 30,65% dos deslocamentos são realizados a pé e de bicicleta, o meio mais saudável tanto para a população quanto para a cidade, motivo pelo qual as atividades relativas às melhorias dos trechos que são feitos a pé, terão prioridade sobre os demais. O transporte motorizado corresponde a 69,35% dos deslocamentos, sendo fator agravante para o congestionamento e a poluição na cidade, motivo pelo qual a área de planejamento de transporte prevê a promoção de qualidade ambiental através da redução da ne-

⁹Ver https://pt.wikipedia.org/wiki/Lista_de_regioes_metropolitanas_do_Brasil_por_populacao, visto em 18 de agosto de 2017.

cessidade de viagens motorizadas, estimulando a adoção do transporte público coletivo.

No ranking de mortes no trânsito, São Paulo é a 13^a cidade do ranking nacional, com 13 mortes para cada 100 mil habitantes, precedida de Porto Alegre com 14,9 e sucedida pelo Rio de Janeiro com 12,9. O líder do ranking é a cidade de Goiânia, com 42,4 mortes para cada 100 mil habitantes. Em São Paulo, as maiores vítimas são os pedestres e os motociclistas, correspondendo a 48%.

O PlanMob 2015 de São Paulo [32] mostra onde estão as estações de monitoramento de poluição e caracteriza a emissão de gases poluentes sob perfis de emissão de gases por tipo de veículo e por hora do dia. A combustão dos veículos libera diversos gases tóxicos na atmosfera; como por exemplo, monóxido de carbono (CO) e hidrocarbonetos (HC). Altos níveis de CO estão associados a diminuição de reflexo, da capacidade de estimar intervalos de tempo, da capacidade de aprendizado; os HC são responsáveis pelo aumento da incidência de câncer de pulmão, provocam irritação no nariz, olhos, pele e aparelho respiratório¹⁰.

Visando a melhoria da mobilidade urbana de São Paulo, o PlanMob 2015 [32] apresenta políticas públicas que envolvam maior rapidez, confiabilidade, regularidade, segurança, conforto e conectividade dos sistemas de transporte. Uma das metas é o melhoramento do sistema de transporte coletivo, através da implantação de novos corredores e faixas exclusivas para ônibus, ampliação de terminais, aumento da velocidade operacional nos principais eixos de transporte coletivo e revisão da rede de linhas de ônibus existente. Também está previsto um estímulo ao uso do sistema cicloviário e a revitalização de parques, visando proteger e recuperar áreas de preservação ambiental, estabelecendo-se “corredores verdes”.

A Região Metropolitana do Rio de Janeiro (RMRJ) é a segunda maior do Brasil¹¹, esta região foi instituída pela lei complementar estadual n°87, de 16 de De-

¹⁰Ver <http://portaldocoracao.com.br/maleficios-a-saude-ocasionados-por-poluentes-atmosfericos/>, visto em 18 de Agosto de 2017.

¹¹Ver https://pt.wikipedia.org/wiki/Lista_de_regiões_metropolitanas_do_Brasil_por_população, visto em 18 de agosto de 2017.

zembro de 1997, e prevê o planejamento e a execução de funções públicas em comum. O relatório da área de transportes da RMRJ se chama Plano Diretor de Transporte Urbano (PDTU). Sendo de particular interesse para a presente tese, foram consideradas duas versões do PDTU para a RMRJ, uma referente ao ano de 2003 e outra referente ao ano de 2013. A seguir, serão mostrados os pontos característicos de cada relatório, e quando possível, indicadores comparativos entre os dois.

O PDTU de 2003 [3] considerou entrevistas domiciliares com 99.310 pessoas de 20 municípios agregados em 17 macrozonas. O PDTU de 2013 [4] subdividiu a RMRJ em 11 macrozonas, um nível de agregação maior do que o anterior. Além de entrevistas e dados complementares obtidos em rodoviárias, estações de trens, metrô e barcas, o PDTU de 2013 considerou contagens volumétricas e pesquisas no cordão externo e interno. Esses dados permitiram a elaboração de matrizes de viagens entre as origens e destinos das zonas de tráfego da RMRJ, evidenciando o motivo e o modo de transporte escolhido.

Em relação a participação no total de viagens de todos os modos de transporte avaliados, o PDTU de 2003 apresentou maiores proporções para viagens realizadas a pé, de bicicleta e de vans. Comparativamente, todos os demais meios de transporte obtiveram um aumento percentual em 2013. Os maiores aumentos foram registrados para viagens de automóvel e ônibus. O gráfico da Figura 2.4 foi montado a partir dos dados correspondentes dos dois relatórios.

Comparativo (%) entre modos de transporte segundo o PDTU de 2003 (Y) e o PDTU de 2013 (X).

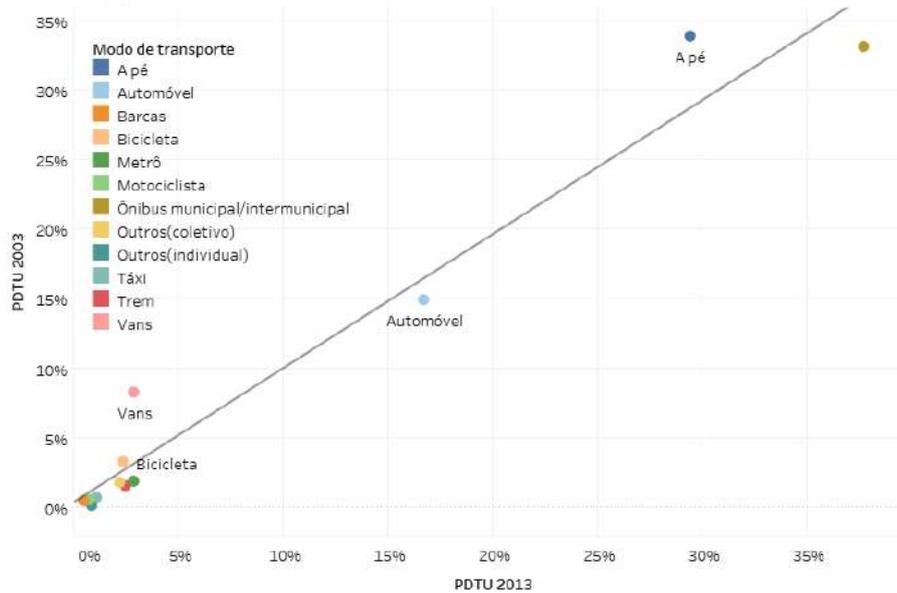


Figura 2.4: Comparativo percentual entre os modos de viagens. Fonte: Elaboração própria a partir de dados do PDTU de 2003 [3] e de 2013 [4].

O reflexo da diminuição das viagens de bicicleta e a pé pode ser visto no aumento da proporção de viagens motorizadas. O PDTU de 2013 [4] apresenta um aumento de 5,3% na proporção de viagens motorizadas. Ou seja, de 62,9%, subiu para 68,2%. Aumentaram basicamente as viagens de táxi, motociclistas e passageiros de automóveis.

O PDTU 2003 [3] considerou as pessoas que não fazem viagem alguma, ou seja a taxa de imobilidade, de onde se vê que 46,6% das pessoas não se movimentavam em 2003. As maiores taxas de imobilidade se verificaram em habitantes com mais de 60 anos. Esse indicador não foi abordado no PDTU 2013.

De forma geral, o PDTU de 2013 [4] buscou a atualização e o enriquecimento do PDTU de 2003 [3] ampliando a quantidade de temas abordados, pode-se destacar alguns:

- A capacidade de passageiros para os trens, metrô, barcas, ônibus;
- O investimento feito no sistema ferroviário e metroviário;

- O investimento feito na construção do arco metropolitano, visando melhorar o fluxo entre Duque de Caxias e Itaguaí¹²;
- A frágil condição das integrações modais, que acabam estimulando a superposição de linhas;
- A infraestrutura inadequada à circulação de portadores de necessidades especiais, a até mesmo de pedestres. Esse item chamou a atenção em diversos terminais;
- A implementação do bilhete único, que possibilitou a reestruturação de linhas de ônibus;
- O efeito *centralizador de demanda* ocasionado pelo crescimento desordenado de novos eixos acoplados a eixos viários já existentes.

Enfim, segundo o próprio PDTU 2013 [4], a frágil integração física operacional entre os modos diminui a eficácia geral do sistema de transportes na RMRJ.

Foram vistas as políticas e os planos de transporte dos dois maiores centros urbanos do mundo e dos dois maiores do Brasil. Cada plano tem abordagens específicas, relativas às questões locais mais latentes.

De forma geral, a organização da mobilidade urbana exerce grande influência na vida das pessoas. Pesquisadores atentos a essa problemática vêm buscando meios alternativos de estudar a mobilidade humana, conforme será visto a seguir.

2.2 Dados de telefonia móvel

O primeiro estudo envolvendo dados de “telefonia” foi realizado com dados de telégrafos em 1949, onde George Zipf [33] descobriu duas leis de potência entre: *a)* quantidade de mensagens trocadas *versus* distância entre um telégrafo e outro, e *b)* soma da quantidade de dinheiro gasto em passagens *versus* distância entre a origem e o destino.

¹²No capítulo de resultados será mostrado o reflexo dessa obra entre Itaguaí e Duque de Caxias.

A partir de então, os registros dos telefones têm sido usados para diversos outros estudos, mas o surgimento dos telefones celulares em quantidade massiva, e de computadores e métodos capazes de lidar eficientemente com esse volume, tornou capaz o surgimento de um grande avanço nessa área [34].

A evolução tecnológica na área de telefonia alavancou mudanças no hábito de comunicação das pessoas. A proporção de pessoas com telefone celular no mundo subiu de 12% no ano 2000 para 96% em 2014 [12]. A tendência multidisciplinar dos estudos de mobilidade mostra que os métodos tradicionais da área de planejamento de transportes podem usufruir de recursos modernos, e que os métodos inovadores das áreas da física e da ciência da computação precisam estar embasados nos tradicionais [35], principalmente na validação [36] das metodologias novas.

A ubiquidade da tecnologia de telefonia móvel vem trazendo à tona os mais diversos usos: sismógrafos distribuídos, sensores de tráfego em auto-estrada, transmissor de imagens médicas [37], detector de comportamentos fora do comum, tais como desastres, acidentes, protestos [38] e até mesmo um contador populacional [39]. Diversos estudos [34] apontam para o uso de dados de telefonia móvel como insumo viável para projetos de pesquisa que podem dar suporte ao planejamento da área transportes.

Dentre os dados gerados pelas operadoras de telefonia, estão os registros de passivos de GPS (Global Positioning System) e de chamadas de telefones celulares, ou CDR (Call Detail Records). Metodologias alternativas vêm sendo desenvolvidas para utilizar esses dados em modelos de transporte [40]. Os dados de CDR são gerados em grande quantidade como subproduto de bilhetagem das operadoras e contêm, dentre outras informações, a localização aproximada do local de realização da chamada telefônica assim como a data e a hora dessa ligação [41]. Apesar das limitações em aferir a escolha modal e as rotas escolhidas, a utilização de CDR permite identificar as principais linhas de desejo da população em

estudo. A seguir serão vistos alguns estudos de mobilidade feitos a partir desses dados.

A partir de 6 semanas de dados de CDR de 2 milhões de pessoas, YANG *et al.* [42] estabeleceram algumas regras espaço-temporais sobre o que caracteriza uma pessoa estar em casa, no trabalho, ou em outro local. A partir dessas regras foi treinado um modelo capaz de extrair as atividades de movimento individual, classificando-as de acordo com essas regras.

Apesar do alto custo em modelar a mobilidade através de entrevistas (coleta *ativa*, conforme visto na Seção 2.1.1), a habilidade de modelar a mobilidade urbana tem grande impacto na qualidade de vida e no funcionamento das cidades. O modelo apresentado por YANG *et al.* [42], não só foi capaz de extrair a mobilidade das pessoas a partir de dados *passivos* de telefonia móvel, como também foi capaz de simular e inferir os motivos mais usuais.

A motivação da viagem revela um aspecto comportamental das pessoas, que pode ser detectado a partir de padrões de frequência de visitas em determinados dias e locais, extraídos dos dados de CDR. Através de algoritmos de *machine learning* e 1 ano dados de chamadas de 80 usuários, LIU *et al.* [43] estabeleceram um modelo capaz de inferir o motivo da viagem. Os motivos considerados foram: (1) casa, (2) trabalho ou escola, (3) outro, (4) visita social, e (5) lazer.

Os dados de CDR fornecem “pegadas digitais” anônimas, revelando padrões de mobilidade humana num nível de detalhamento que talvez não seja possível alcançar através de entrevistas. Em compensação, muitas informações contempladas nas entrevistas não estão disponíveis nos CDR, tais como dados sócio-econômicos, motivo e modo da viagem. Ademais, a localização é aproximada através da localização de uma torre de telefonia móvel e dependente do usuário realizar uma chamada [44].

A partir de 8 bilhões de registros de CDR de 2 milhões de usuários durante 2 meses, ALEXANDER *et al.* [44] estabeleceram uma metodologia capaz de extrair as origens e destinos dos usuários, por motivo e hora do dia. Dois pontos chave

são destacados: a inferência do local de residência dos usuários de telefonia, e o fator de expansão dos usuários à população real. Dos 2 milhões, apenas 335.795 tiveram o seu local de residência descoberto, os quais foram agregados ao nível do setor censitário para gerar um fator de expansão aderente ao censo oficial. Os resultados obtidos foram validados e mostraram boa relação com duas pesquisas oficiais da área de transportes da região. Uma metodologia similar foi adotada na presente tese e será apresentada na seção de metodologia.

Os dados de CDR geralmente demandam por uma grande área de armazenamento e processamento, de tal forma que poucos estudos apresentam um universo amostral que envolva longos períodos. JÄRV *et al.* [5] usaram 1 ano de dados de CDR para compreender a variabilidade periódica no comportamento de mobilidade das pessoas, seja por dia, por mês ou por estação do ano. Essa variabilidade foi medida através dos seguintes *predicados temporais*: (1) número de locais únicos visitados por mês, (2) número de locais únicos visitados por dia, (3) número de locais anualmente visitados, e (4) os 10 locais mais visitados (exceto casa); e *predicados espaciais*: (a) a área em Km^2 em que a pessoa se movimenta por mês, e (b) a área em Km^2 em que a pessoa se movimenta por dia. Os resultados mostraram a alta variação no comportamento de mobilidade entre o verão e o inverno na Estônia.

Os dados de CDR estão vinculados a área de cobertura das torres de telefonia, que geralmente seguem um formato geograficamente irregular, muitas vezes sujeita a acidentes naturais, de forma que a cobertura de uma antena usualmente termina na medida em que começa a de outra¹³.

As principais referências têm adotado a aproximação da área de cobertura das torres através de polígonos de voronoi (ver Figura 2.5 e referência ampla em [45]), que são um tipo de decomposição espacial baseado em divisões equidistantes entre pontos, muito similar a um mosaico.

¹³Dependendo das potências e sobrecargas das antenas, é possível que esse limite seja mais difuso, avançando ou retrocedendo conforme o caso.

A decomposição da área de estudo desta tese via polígonos de voronoi, pode ser vista na Figura 2.6. Mais detalhes da área de estudo serão apresentados no capítulo de conjuntos de dados.

Dado que as delimitações dos polígonos de voronoi, por si só, não encontram correspondência nas delimitações do sistema censitário, o fator de expansão dos usuários de telefonia à população real, é feito pelo uso da relação entre os polígonos de voronoi, as unidades de um *grid* regular e os setores censitários, ver Figura 2.7 mais à frente. A partir dos dados de CDR, toda a informação sobre os padrões de mobilidade dos usuários é inferida por de níveis de agregação entre um quadrado do *grid* e outro [46].

Quanto maior a densidade de antenas, maior a precisão em delimitar a localidade do indivíduo. A partir de 10 meses de dados de CDR de Portugal e 5 meses de dados de CDR da França, DEVILLE *et al.* [47] inferiram a população residente nos setores censitários a partir da quantidade de usuários que registraram chamadas nas torres de telefonia. Basicamente, essa inferência se deu em três etapas conforme a seguir.

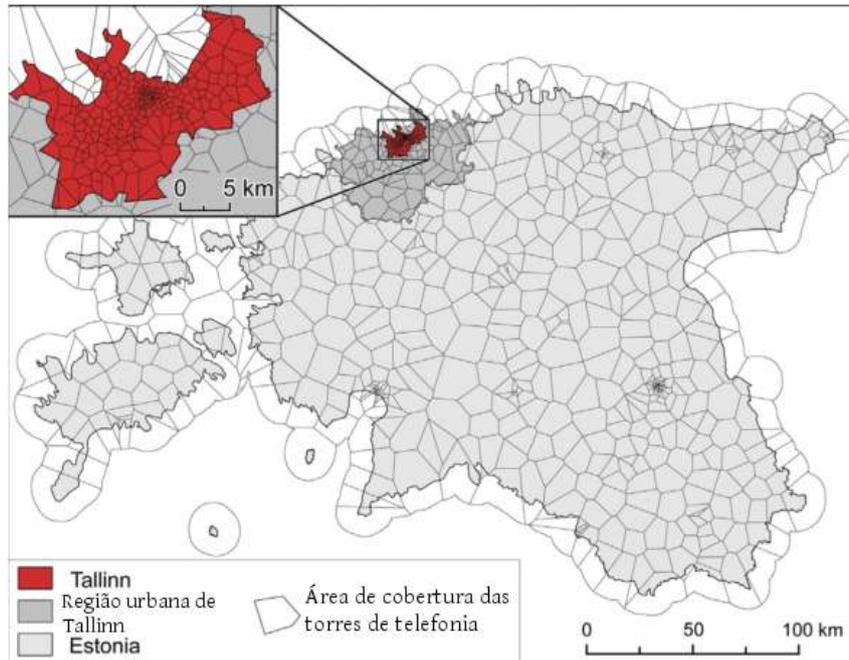


Figura 2.5: Exemplo de área de cobertura das torres de telefonia com o uso de polígonos de voronoi. Fonte: Figura 1 de [5], foram feitas as traduções ao português.

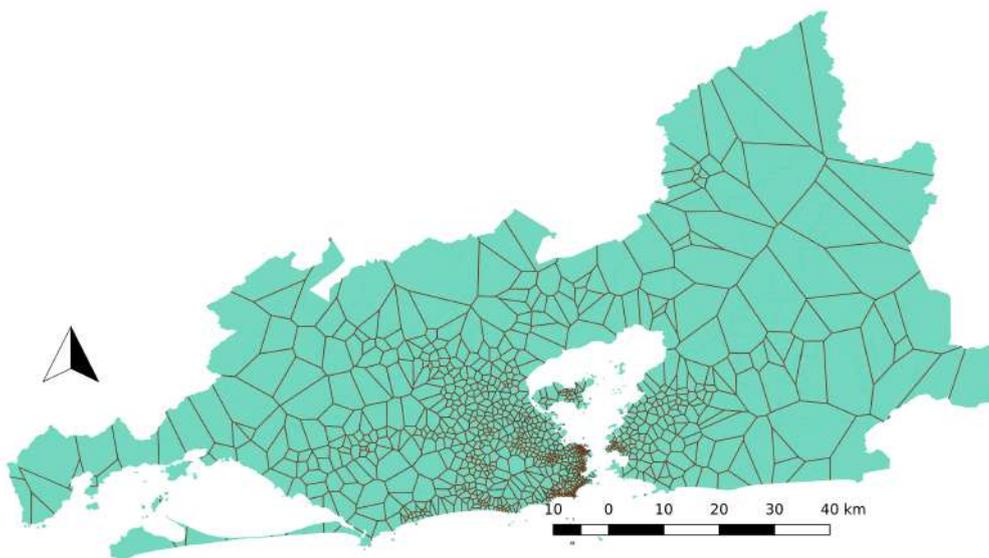


Figura 2.6: Área de cobertura das torres de telefonia da região de estudo da presente tese, através de polígonos de voronoi. Fonte: Elaboração própria.

A densidade de usuários sob a área de cobertura de uma torre de telefonia j , é dada por DEVILLE *et al.* [47] :

$$\sigma_{V_j} = \frac{T_j}{A_{V_j}} \quad (2.6)$$

onde: T_j é a quantidade de usuários que tiveram atividades registradas no período entre 20h e 7h (considerando que o usuário estava em casa neste período),

V_j é o polígono de voronoi associado à torre j ,

A_{V_j} é a área do polígono V_j .

Resumidamente, usuários de telefonia detectados no polígono de voronoi divididos pela área do mesmo polígono.

A estimativa da densidade de usuários no setor censitário c_i , é dada por:

$$\sigma_{c_i} = \frac{1}{A_{c_i}} \sum_{V_j} \sigma_{V_j} A_{(c_i \cap V_j)} \quad (2.7)$$

onde: A_{c_i} é a área do setor censitário c_i . Ou seja, a densidade de usuários sob a área da torre j é transferida à área do setor c .

Para escalar a densidade de usuários à densidade populacional, foram treinados dois coeficientes, sendo um o fator de escala α e o outro um expoente β a partir da população oficial do censo, da seguinte forma:

$$\rho_c = \alpha \sigma_c^\beta \quad (2.8)$$

onde: $\rho_c = [\rho_{c_1}, \rho_{c_2}, \dots, \rho_{c_n}]$, é a densidade população oficial do censo no setor c , e $\sigma_c = [\sigma_{c_1}, \sigma_{c_2}, \dots, \sigma_{c_n}]$, é a densidade de usuários detectados como residentes no mesmo setor c .

Uma validação da metodologia de DEVILLE *et al.* [47] foi feita por eles mesmos a partir do recurso LandScan¹⁴ [48], que consiste basicamente em inferir a densidade demográfica em qualquer local do globo a partir de imagens de saté-

¹⁴Ver <http://web.ornl.gov/sci/landscan/index.shtml>, visto em 30 de agosto de 2017.

lite sob o formato de um grid regular. No caso específico dessa validação, cada quadrado do grid possuía uma área de $100m^2$.

De forma geral, os resultados da metodologia de inferência populacional a partir de dados de celular [47] obtiveram melhor precisão em áreas de maior densidade populacional, superestimando a população em áreas menos densas. Globalmente, a população inferida através de imagens de satélite (LandScan) foi mais precisa. Um exemplo ilustrativo comparando as delimitações entre polígonos de voronoi e grid regular pode ser visto na Figura 2.7.

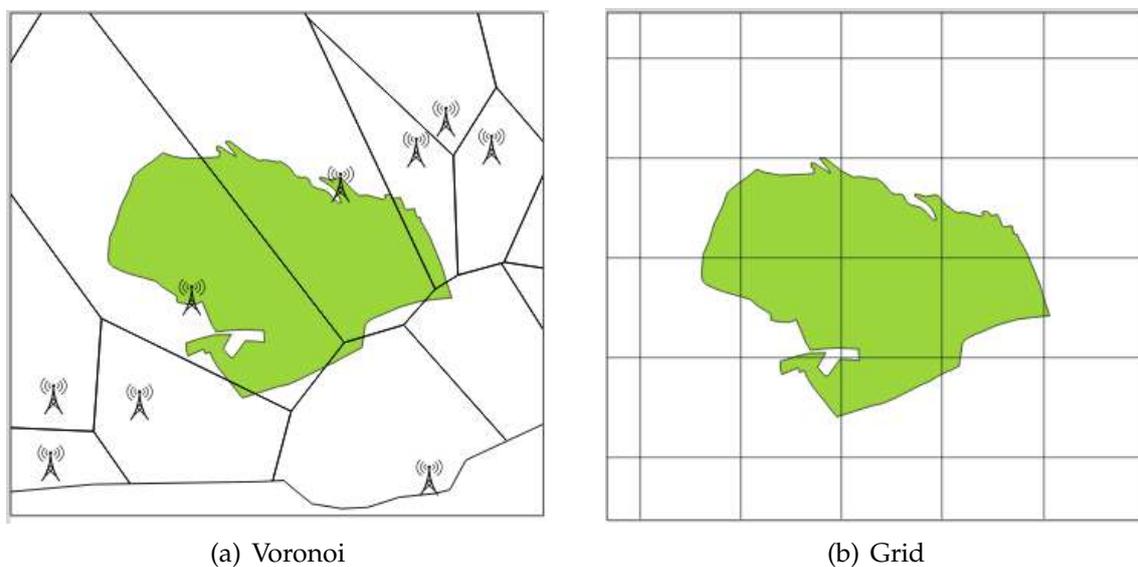


Figura 2.7: Um mesmo subdistrito é destacado em verde (Rocinha), mostrando a sobreposição entre a área de cobertura de torres de telefonia e um grid regular, imagens (a) e (b) respectivamente.

Em mais uma referência de subdivisão geográfica com grid regular, WANG *et al.* [49] apresentaram uma metodologia de particionamento capaz de inferir zonas geográficas, que basicamente são zonas de tráfego representadas por agregações de um grid regular, obtidos a partir do agrupamento de dados de GPS de táxi. A ideia básica é que os percursos mais usuais dos taxistas, medidos pelos locais de embarque e desembarque dos passageiros, formem zonas de tráfego com características espaciais semelhantes.

Sob o ponto de vista da área de cobertura das torres de telefonia, é possível que ocorram imprecisões sobre a real posição geográfica dos usuários na leitura

dos dados de CDR. SCHNEIDER *et al.* [50] estabeleceram algumas regras para amenizar tais circunstâncias: (1) registros de torres com menos de 0,5% de uso durante todo o período de estudo são eliminados, (2) registros que oscilam três vezes no mesmo dia entre torres vizinhas são mesclados, (3) dias com atividade abaixo do normal são descartados segundo critérios pré-estabelecidos.

Diversas metodologias que caracterizam, simulam e validam a mobilidade humana a partir de dados de CDR continuam a surgir. JIANG *et al.* [51] apresentam um fluxo abrangente para avaliar a mobilidade: (1) armazenamento dos dados brutos de CDR, (1.1) extração de comportamento, se o usuário ficou ou apenas passou pelo local, (1.2) detecção do local de residência, (1.3) exclusão de usuários que não possuem um volume pré-definido de atividade; (2) armazenamento dos dados do censo, (2.1) elaboração do particionamento geográfico de estudo; (3) extração das redes individuais de mobilidade dos usuários, *i.e.*, casa-trabalho, casa-trabalho-escola, etc; (4) cálculo dos fatores de expansão à população real; e (5) agregação e visualização dos padrões de mobilidade da população na região de estudo. Uma metodologia similar foi adotada nesta tese e será apresentada nos capítulos oportunos.

2.2.1 Matriz O-D

O posicionamento da matriz O-D em relação ao modelo de quatro etapas já foi visto na seção de distribuição de viagens do FSM (Seção 2.1.1). A inferência de viagens a partir de dados de CDR (ou de GPS) se dá em uma ordem diferente do modelo tradicional, pois as viagens entre as *origens* e *destinos* são descobertas antes dos polos *produtores-atratores*. Isso ocorre porque ao se inferir uma viagem a partir de duas chamadas telefônicas consecutivas, a informação de local de origem e de local de destino já é automaticamente revelada. A seguir serão vistas referências específicas de inferência de matriz O-D a partir de dados de celular.

De forma semelhante à área de transportes, para se inferir mobilidade a partir de dados de celular, é necessário um conhecimento prévio da área de estudo. É

preciso que exista alguma referência para a definição do que venham a ser origens e destinos. Essas referências são em grande parte embasadas no censo demográfico, na área de cobertura das torres de telefonia, e na definição de um grid regular [7], [51]. Um exemplo dessas delimitações já foi mostrado nas Figuras 2.5 e 2.7.

Por mais precisa que seja a inferência de uma matriz O-D para a área de transportes, ela pode estar sujeita a imprevistos. Por essa razão, o levantamento das entrevistas domiciliares precisa ser extenso, ao ponto de obter uma “fotografia” de mobilidade que expresse um dia útil comum. Ou seja, a natureza do trânsito é dinâmica e demanda por metodologias que a considerem da mesma forma. MOREIRA-MATIAS *et al.* [52] apresentaram uma metodologia para estimativa da mobilidade urbana em tempo real, através de dados contínuos de GPS de táxi.

Visto que o dado de GPS é contínuo, é necessária uma discretização espaço-temporal. A principal contribuição de MOREIRA-MATIAS *et al.* [52] é uma metodologia capaz de realizar essa discretização de forma incremental sob carga contínua de dados de GPS, mantendo as estatísticas de interesse da área de transporte atualizadas em relação aos pares O-D da área de estudo.

Mesmo que o usuário de telefonia não realize e nem receba chamada alguma, e também não disponha de um aplicativo de GPS operacional, as antenas de telefonia registram o aparelho celular, “alocando-o” em uma área de cobertura, possibilitando que ele realize ou receba chamadas. Essa alocação através das antenas da rede de telefonia permite acompanhar o usuário nos percursos realizados.

Assumindo que cada celular pode representar um veículo, e que um conjunto de torres de telefonia (do tipo GSM) pode representar uma zona de tráfego, CACERES *et al.* [6] apresentaram uma metodologia para inferência de matrizes O-D através de registros de *passagem* dos celulares entre as antenas, ver Figura 2.8, onde aparece um veículo saindo da área A1 e entrando na área A2. Observe que no sistema de telefonia GSM, a cobertura das antenas é inferida por hexágonos, não por voronoi.

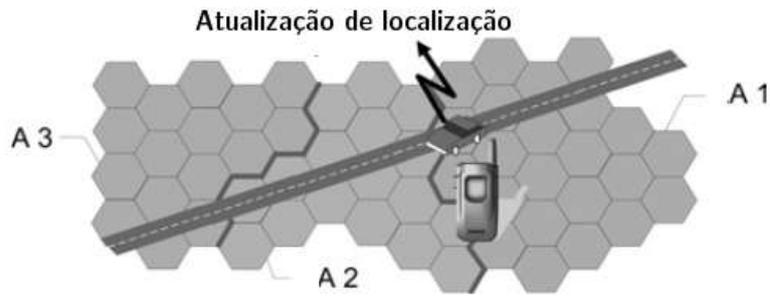


Figura 2.8: Mudança de localização através de uma rede de telefonia GSM, cada hexágono representa uma antena. Fonte CACERES *et al.* [6], traduzido ao português.

A partir de bilhões de dados de CDR, de dados do censo demográfico, e de mapas públicos, TOOLE *et al.* [7] apresentaram uma metodologia que, dentre outras contribuições, foi capaz de inferir os *fluxos transientes* dos usuários de telefonia móvel e classificar as principais vias de 5 cidades (Boston, Baía de São Francisco, Rio de Janeiro, Lisboa e cidade do Porto) entre 4 categorias de uso. A metodologia contempla a criação de matrizes de origem e destino transiente entre os subdistritos censitários, com fator de expansão à população real, inferência de rotas possíveis, e culmina na visualização interativa do VOC (*Volume-Over-Capacity*) das principais vias do eixo rodoviário.

O fluxo transiente, ou intermediário, ocorre quando o usuário sai de um ponto a outro da cidade, mas esses pontos não necessariamente significam as origens e destinos finais respectivamente, vide Figura 2.9; observa-se que o usuário saiu do ponto A em direção ao ponto D, contudo ele realizou uma chamada no ponto B e outra no ponto C, revelando o fluxo transiente BC.

Algumas pesquisas vêm explorando a possibilidade de cruzamento de dados para obter metodologias mais robustas. Através de dados de telefonia móvel e contagem de veículos em rodovias, IQBAL *et al.* [53] propuseram uma metodologia capaz de inferir matrizes O-D transientes entre interconexões de tráfego. O estudo foi possível a partir de um mês de dados de CDR de 6,9 milhões de pessoas de Dhaka, Bangladesh, e três dias de vídeos de diversos pontos da cidade, contendo informações sobre a contagem de tráfego.

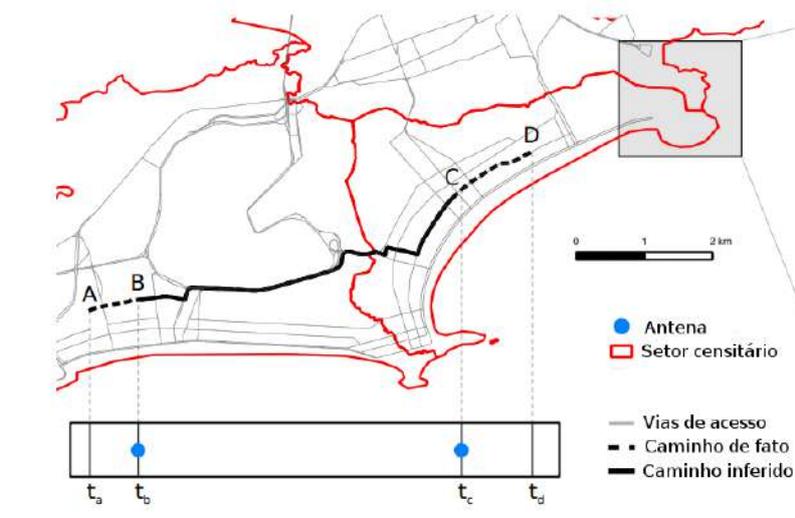


Figura 2.9: É visto um fluxo transiente entre B e C, quando na verdade o usuário partiu de A até D. Fonte TOOLE *et al.* [7], traduzido ao português.

As viagens obtidas através dos dados de CDR obtiveram um fator de expansão para se equiparar à contagem real de veículos, coletada através de filmagens [53]. Para diminuir a detecção de *falsas viagens*, foi observado um intervalo de tempo Δt entre uma chamada e outra. Ou seja, viagens detectadas através de chamadas consecutivas, que ocorreram num intervalo menor do que 10min e maior do que 1h ($10min > \Delta t > 1h$) foram descartadas. Essa parametrização é muito específica do local e do foco do estudo. Para regiões maiores é possível esses limites sejam ajustados com valores diferentes. Parâmetros similares foram adotados nesta tese.

Sob o contexto de crescimento urbano no Senegal, onde a população urbana subiu de 19% para 39% em cerca de 50 anos, DEMISSIE *et al.* [54] aplicaram uma metodologia de inferência de matriz O-D a partir de dados de CDR. O caso de estudo foi focado no Senegal, contudo a metodologia pode ser aplicada em qualquer país em desenvolvimento, pois aborda aspectos comuns a esses países, tais como a predominância de sistemas de transporte informal, geralmente um sinal de que as políticas públicas não acompanharam a velocidade da demanda da população.

Por uma medida extra de proteção de privacidade¹⁵, a companhia telefônica que forneceu os dados de CDR para DEMISSIE *et al.* [54], não forneceu as coordenadas exatas das antenas, dando em troca uma localidade (Lat/Long) aproximada do local real de onde ocorreu o registro. Foram fornecidos dados de 300 mil usuários randômicos durante 13 dias, obedecendo os seguintes critérios: *a*) usuários com registros em no mínimo 75% dos dias, e *b*) usuários com no máximo 1000 registros por semana. Por convenção, foi assumido que celulares com mais de 1000 registros por semana eram máquinas ou eram aparelhos compartilhados.

Os dados mostraram claramente que o serviço de transporte público da região não atendia toda a demanda da população. DEMISSIE *et al.* [54] sugeriram a introdução de novas rotas de ônibus e uma melhoria nas rotas existentes. As novas rotas sugeridas levaram em conta principalmente a quantidade de viagens inferidas entre as origens e os destinos.

Enfim, foram vistas diversas referências que usaram dados de telefonia móvel especificamente para inferências de matriz O-D, um dos focos desta tese.

2.3 Análise e planejamento de eventos

Além dos dados de entrevistas sobre a mobilidade quotidiana da população usados pelos engenheiros de transporte, foi introduzida a possibilidade de uso dos dados de telefonia móvel, essa parece ser já uma realidade segundo revistas específicas dessa área. Contudo é cada vez mais comum um tipo diverso de demanda de mobilidade, que se enquadra com propriedades afins na área de turismo, trata-se do planejamento de eventos.

A seguir serão apresentados alguns conceitos de eventos, através de um autor chave para pesquisas nessa área, sendo a primeira abordagem feita em 2008 (GETZ [55]), e um novo artigo em 2016 (GETZ e PAGE [8]) atualizando o mesmo tema abordado em 2008¹⁶.

¹⁵Além do mascaramento do número do celular.

¹⁶Consultado a partir do portal da capes, na base *Web of Science*, em 16 de setembro de 2017.

Existem vários tipos de eventos, alguns podem ter origem em determinados costumes comunitários locais ou regionais, que pela sua capacidade de atrair o público, acabam por tornar-se atividades profissionais desempenhadas pelo setor de turismo. Geralmente este tipo de evento está ligado a uma identidade cultural, relacionada a tradições e peregrinações religiosas, festivais gastronômicos, carnavais e outras celebrações.

Existem eventos que podem ser planejados com a finalidade de fomentar o comércio e promover economicamente uma determinada região. Geralmente esse tipo de evento está ligado à área de lazer e entretenimento. Ou seja, lazer desenvolvido através de planejamento e pesquisa.

Embora haja uma preocupação particular de que o turismo corrompa a identidade cultural dos eventos, ele é a força empreendedora que conduz e motiva os esforços nessa área. Há, inclusive, argumentações contrárias, em que o turismo ajudaria a preservar as tradições através de festivais e outras celebrações culturais.

Basicamente, os eventos podem ser classificados quanto ao tipo e a dimensão ou abrangência, de forma que uma mesma região pode ser palco de diversos eventos seguindo uma determinada sazonalidade, ou uma propensão da população em ser atraída para determinada finalidade, ver Figuras 2.10 e 2.11, segundo [8].

A pirâmide de eventos (Fig. 2.11) mostra os tipos de evento segundo uma abrangência e uma rotina. A partir da base, estão os eventos mais comuns de ocorrer, e ao mesmo tempo os que atraem um público mais próximo e regional; no topo estão os eventos mais raros, que acabem por atrair pessoas do mundo inteiro, ou seja, os megaeventos.

Recentemente, diversos trabalhos têm sido realizados com foco em megaeventos. Megaeventos, são eventos de nível global que exigem um planejamento de médio/longo prazo e demandam a criação ou expansão de infraestruturas de sistemas de transportes e maiores intervenções urbanísticas nas cidades sedes [56].



Figura 2.10: Tipos de eventos segundo a perspectiva de turismo. Fonte: Elaboração própria, a partir de GETZ e PAGE [8].



Figura 2.11: Tipos de eventos segundo escala ou abrangência turística. Fonte: Elaboração própria, a partir de GETZ e PAGE [8].

Há pelo menos três tipos de eventos que podem ser considerados como megaeventos: Copa do Mundo de Futebol, Jogos Olímpicos e Exposições Mundiais [56] [57].

O Rio de Janeiro foi palco de alguns megaeventos recentemente, a Copa do mundo em 2014 e as Olimpíadas em 2016. Às vésperas desses eventos, QUIN-

TELLA [58] discorreu sobre questões de planejamento e segurança regionais, focando nos desafios da área de segurança pública para suportar eventos de dimensões internacionais no Rio de Janeiro.

Abaixo dos megaeventos estão os eventos de grande porte, que atraem visitantes de outras cidades ou países, mas a grande parcela de participantes é de localidades próximas e não requerem grandes alterações de infraestrutura. São exemplos de eventos de grande porte: festas populares, grandes festivais de música, e eventos esportivos. Megaeventos acontecem raramente na mesma cidade. Eventos de grande porte são mais comuns e pressionam os sistemas de transportes e o sistema viário existente por um curto período de tempo (algumas horas ou poucos dias) e demandam pequenas intervenções de tráfego e programações especiais dos sistemas existentes de transporte público.

No caso de eventos fora da rotina, há uma dificuldade na obtenção de dados para realização de previsões de demanda e planejamento de transportes. Além da população local, a presença de turistas de diferentes localidades altera a dinâmica dos sistemas de mobilidade e do transporte coletivo em um curto período de tempo, tornando o planejamento desses sistemas uma tarefa árdua e imprecisa.

A partir de dados de CDR foram feitos diversos estudos de mobilidade focados em eventos, os objetivos foram desde a monitoração de tráfego nos grandes centros urbanos até a detecção de conflitos de rua. Serão mostradas a seguir algumas dessas contribuições.

2.3.1 Uso de dados de CDR para análise de eventos

Existem alguns estudos feitos a partir de dados de CDR que focaram em eventos na RMRJ. JOSÉ [18] analisou a origem das pessoas que foram ao Reveillon em Copacabana e num tradicional bloco de carnaval do Rio de Janeiro. O local de residência dos usuários de telefonia foi presumido a partir do padrão de chamadas telefônicas. Uma abordagem similar foi adotada no presente estudo e será detalhada no capítulo de estimativas de matrizes de origem e destino.

Sob um enfoque na carga de trabalho das torres de telefonia por conta de megaeventos e eventos de grande porte, XAVIER *et al.* delimitaram os locais e os intervalos de tempo de maior aglomeração de pessoas na Copa do Mundo de 2014 [17], e no campeonato Brasileiro [59] [60], dentre outros eventos. Uma breve descrição dessas dessas caracterizações será apresentada a seguir.

A partir da contagem de chamadas telefônicas registradas pelas antenas situadas na área de cobertura do estádio olímpico Nilton Santos (*Engenhão*), XAVIER *et al.* [59] identificaram dois padrões bem definidos de aglomeração de pessoas: *a)* um para o dia em que houve jogo, e *b)* outro para o dia em que não houve evento. Observou-se que na decorrência do jogo, os torcedores (*cerca de 30 mil pessoas por jogo*) tendem ligar com maior frequência do que os não-torcedores.

Em continuidade à pesquisa de 2012 [59], XAVIER *et al.* [60] analisaram novas métricas em relação às chamadas telefônicas: *a)* duração, *b)* intervalo entre chamadas antes do evento, e *c)* intervalo entre chamadas após o evento. Essas métricas possibilitaram a descoberta sobre onde as pessoas estavam antes do evento e para onde foram depois. O evento escolhido foi o Reveillon de 2012, analisado de forma simultânea em três capitais brasileiras: Belo Horizonte, Recife e Salvador. As antenas atenderam 100 mil, 10 mil e 1 milhão de pessoas respectivamente, no período compreendido entre 21h45 e 2h30. Novamente os resultados foram consistentes em mostrar os momentos de maior carga nos sistemas de telefonia dada pelas concentrações de pessoas.

Sob a mesma problemática abordada em [59] e [60], XAVIER [17] apresentou uma dissertação de mestrado, sob forma mais abrangente no que tange aos diferentes tipos de eventos, desde a Copa do Mundo de 2014, festas de Reveillon e shows até corridas de automobilismo. Como resultado foi apresentada uma ferramenta própria para análise de concentração de recursos da rede de telefonia móvel sob o estado anômalo dos grandes eventos, acarretado pela dinâmica de mobilidade fora do comum nesses megaeventos e eventos de grande porte.

O tema de pesquisa “*analisar a mobilidade urbana em eventos a partir de dados de telefonia móvel*” vem sendo abordado em outros países. A seguir será apresentado um breve resumo sobre pesquisas realizadas sob áreas de estudo localizadas na Itália [61] e no Senegal [38].

CALABRESE *et al.* [61] apresentaram um sistema de monitoramento urbano em tempo real numa plataforma desenvolvida pela Telecom Itália, especificamente para a cidade de Roma. Foram explicadas as limitações dos sensores de GPS e as vantagens de se usar os aparelhos celulares como sensores de mobilidade, dado principalmente pela sua ubiquidade. A área de estudo se aplica à uma região de 100 km² de Roma, levando-se em conta 27 mil chamadas em um dia útil. O sistema apresentado opera em conjunto: dados de celular, dados de GPS de ônibus e dados de GPS de Táxi. Como resultado, foi possível observar a cidade de Roma como um organismo vivo, com seus pontos de atração, onde os turistas estavam se reunindo, se o transporte público estava presente onde as pessoas estavam, e para onde os congestionamentos estavam se encaminhando.

TRESTIAN *et al.* [38] apresentaram um estudo de eventos a partir de CDR de grande abrangência temporal no Senegal, foi analisado um ano de dados entre 1º de Janeiro e 31 de Dezembro de 2013, sob uma área aproximada de 197 mil km² e uma população estimada de 13,5 milhões de pessoas. Dos três conjuntos de dados cedidos pela operadora de telefonia Orange, chegando a média de 11 milhões de chamadas por dia, o mais detalhado considerou 300 mil usuários randômicos a cada duas semanas, atendidos por 1666 antenas. A partir de uma análise estatística, os autores observaram uma correlação linear entre população, antenas e chamadas telefônicas. Tanto grandes eventos religiosos quanto protestos populares, foram observados através dos dados de CDR. Dado que a religião predominante no Senegal é a muçulmana, pode-se observar uma mudança nos hábitos da população durante o Ramadã, viu-se que as pessoas ligam e falam mais ao celular durante a noite e menos durante a parte da manhã, apenas durante o mês do Ramadã. No dia 1º de fevereiro de 2013 foi reportado ataque do

movimento Casamance contras as forças que estavam em Kafoutine, quatro foram mortos. O reflexo pôde ser visto nos dados de CDR do dia posterior através de um acréscimo significativo, tanto na quantidade de chamadas quanto na sua duração, tudo capturado por apenas uma antena de celular em Kafoutine.

Os trabalhos apresentados obtiveram sucesso em monitorações em tempo real, em avaliação de carga de trabalho dos sistemas de telefonia, em detecção de grandes eventos e até mesmo conflitos, de forma geral.

2.4 Modelo de radiação

Conforme visto na Seção 2.1.1, o modelo de quatro etapas tem como pré-requisito a definição de alguns parâmetros, geralmente ligados à características sócio-econômicas da região de estudo, tais como oferta de empregos, renda, quantidade de veículos, etc. O modelo de radiação é proposto por SIMINI *et al.* [9] como uma alternativa ao modelo gravitacional, contudo sem a necessidade de conhecimento prévio da região de estudo, e consequentemente sem necessidade de ajuste de parâmetros. Os únicos requisitos são: a distância entre um ponto e outro, a população residente na cidade/subdistrito de origem e de destino, e a população residente no raio a partir da origem até o destino.

O modelo de radiação parte do princípio que a quantidade de postos de trabalho de uma determinada região está ligada à própria população. Esse princípio, assim como diversos outros, está embasado na lei de potência (*Power-law*), por exemplo (ver BETTENCOURT *et al.* [62]), a população $N(t)$, tomada como medida de uma cidade no instante de tempo t , segue uma lei de potência da seguinte forma:

$$Y(t) = Y_0 N(t)^\beta \quad (2.9)$$

Y denota uma medida de atividade social de interesse, ou seja, renda, registro de patentes, **ofertas de emprego**, etc¹⁷. Y_0 é uma constante de normalização. O

¹⁷Não é qualquer atividade social que segue uma lei de potência, para maior compreensão, ler BETTENCOURT *et al.* [62] e ver a apre-

expoente, β , reflete a dinâmica geral do sistema urbano. A equação 2.9 é similar ao modelo de geração de viagens.

A formulação do modelo de radiação faz analogia com o processo de emissão e absorção de radiação, já conhecido pela área da física, contudo quando aplicado à mobilidade humana, as pessoas tomam o lugar de “partículas”, que são emitidas de um ponto e absorvidas em outro. Segundo o modelo de radiação, o fluxo médio de viagens T_{ij} , entre uma origem i e um destino j , pode ser inferido pelo modelo de radiação, basicamente como :

$$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \quad (2.10)$$

onde m_i e n_j denotam a população da região de origem i e de destino j , respectivamente; s_{ij} denota a população que vive no raio r_{ij} , centrado em i , excetuando-se a população residente nos próprios locais m_i e n_j , conforme pode ser visto na Figura 2.12.

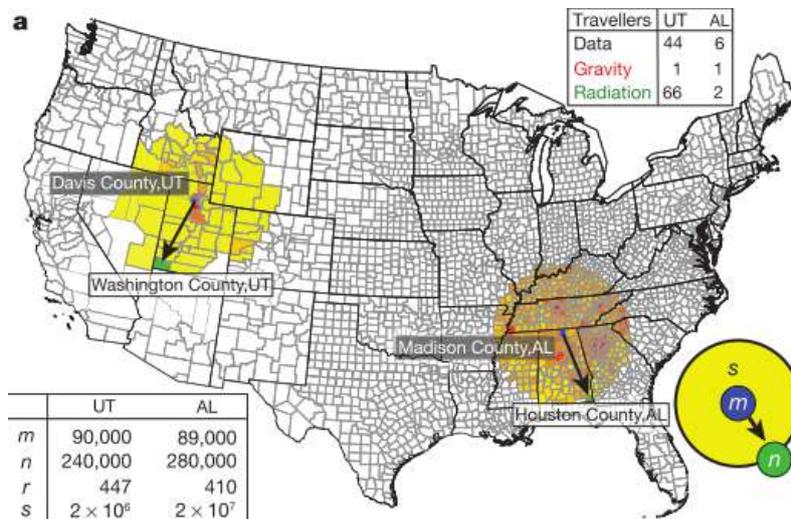


Figura 2.12: População s considerada no modelo de radiação, excetuando-se os residentes de m e n . Fonte SIMINI *et al.* [9].

Alguns autores analisaram comparativamente os resultados obtidos pelos modelos de radiação e de gravitação. Segundo WESOLOWSKI *et al.* [63], os resultados de Geoffrey West no TED Talks https://www.ted.com/talks/geoffrey_west_the_surprising_math_of_cities_and_corporations, visto em 9/12/2017.

delos de mobilidade foram criados a partir de dados de países desenvolvidos, e esse viés deixa os modelos não aderentes aos países sub-desenvolvidos. Foram analisados os os dados de CDR de aproximadamente 15 milhões de usuários anônimos do Quênia durante um ano. Levando-se em conta os próprios dados de chamadas sucessivas como “deslocamentos observados”, foram configurados os modelos de gravitação e de radiação. Ambos falharam, seja superestimando a área de abrangência das viagens, ou capturando erroneamente a mobilidade dos meios rurais aos urbanos.

MASUCCI *et al.* [64] compararam os modelos de gravitação e radiação com dados do Reino Unido. O modelo de radiação apresentou melhores resultados considerando longas distâncias entre origens e destinos. Sob distâncias mais curtas, o modelo gravitacional apresentou maior precisão. Enfim, em linhas gerais o modelo de radiação foi tido como promissor e o gravitacional como mais estável.

2.5 Tópicos conclusivos

Nesta revisão foram vistas referências próprias da área de transporte, assim como relatórios técnicos de alguns dos maiores centros urbanos do mundo. Foram vistas referências sobre o sistema de telefonia móvel, e sobre como as chamadas telefônicas podem ser usadas para inferir questões de transporte.

Observa-se a clara tendência de urbanização em diversos países, inclusive no Brasil. Essa tendência vem acompanhada de diversos desafios em “gerir” regiões cada vez mais densamente povoadas. Esses desafios ficam mais salientes ao se observar a lei federal 12.587 de 3 de janeiro de 2012, que determina que municípios com mais de 20 mil habitantes elaborem um plano de mobilidade urbana. Observe que dos 5.570 municípios brasileiros¹⁸, 1.746 possuem mais de 20 mil habitantes. Sabe-se que apenas alguns poucos centros urbanos possuem planos de mobilidade.

¹⁸Segundo o censo do IBGE de 2015.

Os dados de telefonia celular têm muito a oferecer para auxiliar nos desafios de planejamento de infraestrutura dos grandes centros urbanos, já que: 1) as metodologias envolvendo esses dados contam com a participação *indireta* e abrangente da população; 2) os modelos de mobilidade estabelecidos podem ser atualizados na mesma velocidade em que se obtenham os dados da rede de telefonia; e 3) os dados de telefonia permitem um planejamento numa escala mais ampla do que o método de entrevistas individuais.

O tema de “uso de dados de telefonia para inferir mobilidade” ainda é recente no país, e merece mais pesquisas, dada a possibilidade de benefícios à sociedade, e principalmente aos grandes centros urbanos. Espera-se que esta pesquisa seja uma “abertura” para diversas novas pesquisas.

As referências da área de turismo elucidaram alguns conceitos sobre “eventos”. O estudo de mobilidade da população por conta de eventos extraordinários também parece ter grande potencial de uso, já que torna-se possível a inferência da origem das pessoas nos eventos, revelando até mesmo padrões de comportamento sazonais [5].

Enfim, os dados de telefonia têm muito a oferecer ao planejamento de transportes, contudo não se vê uma “ruptura”, que justifique uma substituição da metodologia tradicional de entrevistas e pesquisas domiciliares. Um indivíduo que vai todos os dias a uma universidade, por exemplo, pode ser um aluno ou um professor¹⁹. Essa definição é de grande importância, pois irá revelar o motivo do deslocamento: se é uma viagem para estudo (HBS), ou uma viagem para trabalho (HBW). Ou seja, o dado de CDR é rico e permite observar a mobilidade em larga escala, contudo pode ser arriscado esperar nele uma resposta para questões mais específicas, como por exemplo o motivo da viagem.

¹⁹Ou funcionário, ou servidor.

Capítulo 3

O conjunto de dados

“Ninguém o pode aconselhar ou ajudar, ninguém. Não há senão um caminho. Procure entrar em si mesmo. Investigue o motivo que o manda escrever; examine se estende suas raízes pelos recantos mais profundos da alma; confesse a si mesmo: morreria, se lhe fosse vedado escrever?”

— Rainer Maria Rilke, Cartas a um jovem poeta.

Conteúdo

3.1 Os dados de CDR	50
3.2 Os dados do censo	56
3.3 Os dados do PDTU	59

O desenvolvimento desta pesquisa só foi possível devido ao acesso, organização, manipulação e cruzamento de dados geo-referenciados delimitados à área de estudo. O geo-referenciamento foi um ponto crucial para alavancar a metodologia.

A adoção da tecnologia de telefonia móvel no Brasil segue a tendência mundial. A Tabela 3.1 mostra o percentual da população com telefone celular, por regiões no Brasil. Em 2005 o percentual da população brasileira que possuía acesso ao telefone celular era de 36,6%, chegando a 78,3% em 2015. Em nível regional, a utilização de telefone celular chega a 82,6% nas regiões Sul e Sudeste e 86,9% na região Centro-Oeste. Em grandes centros urbanos, este percentual atinge mais de 90%. Segundo o site Teleco¹, em 2016 havia no Brasil 1,18 aparelhos de celular por habitante.

Ano	Brasil	Norte	Nordeste	Sudeste	Sul	Centro-Oeste
2005	36,6	26,4	23,9	40,9	47,5	47,5
2008	53,7	43,9	41,2	58,6	62,7	64,3
2013	75,2	66,7	66,1	79,5	79,8	83,8
2015	78,3	68,6	69,6	82,6	82,8	86,9

Tabela 3.1: Percentual de pessoas com telefone móvel celular para uso pessoal, na população de 10 anos ou mais de idade (%) por grandes regiões. Fonte: PNAD IBGE – Pesquisa Nacional por Amostra de Domicílios.

A área de estudo compreende a Região Metropolitana do Rio de Janeiro (RMRJ), ver Figura 3.1. Trata-se da segunda maior área urbana do Brasil.

¹<http://www.teleco.com.br/estatis.asp>, visto em agosto de 2017.

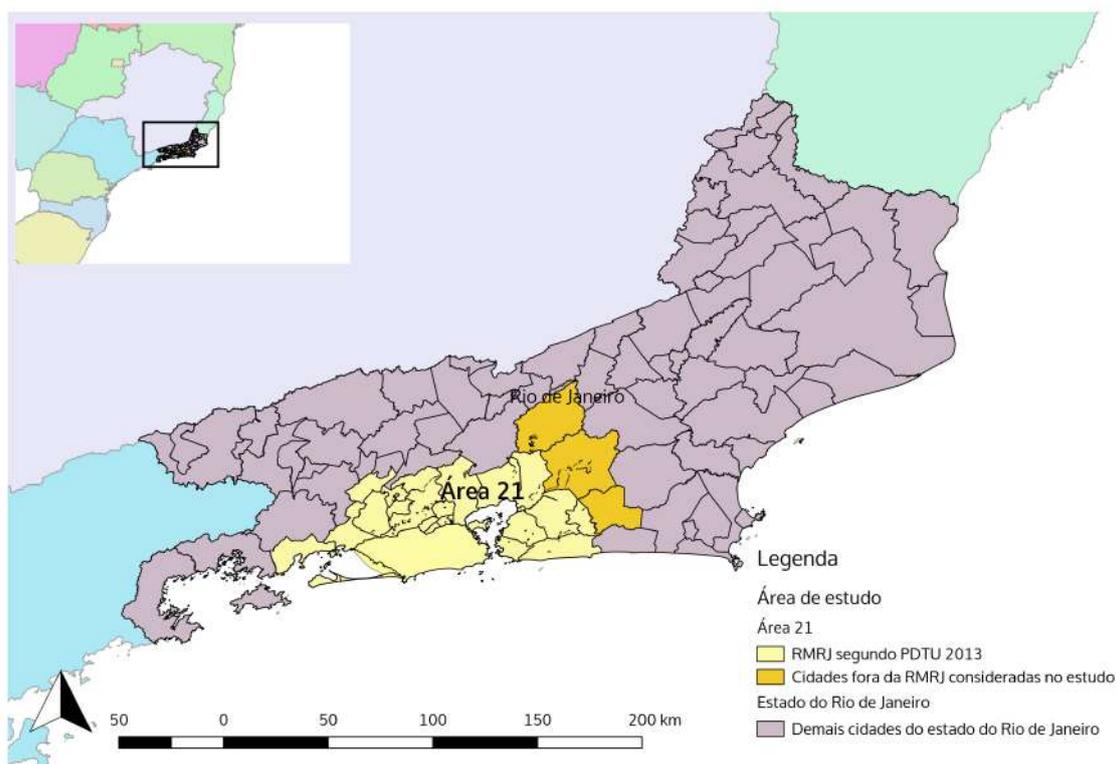


Figura 3.1: Área de estudo.

Segundo dados fornecidos no último relatório da Secretaria de Transporte do Estado do Rio de Janeiro (PDTU 2013 [4]), essa área territorial se estende por 5.326 km², e possui 11.835.708 habitantes distribuídos em 19 municípios, sendo a cidade do Rio de Janeiro, com 6.323.037 habitantes em 1.200 km², a mais importante.

Os dados de chamadas telefônicas foram coletados a partir das antenas localizadas no código de área 21², que corresponde à Região Metropolitana do Rio de Janeiro incluindo-se os municípios de Teresópolis, Mangaratiba, Rio Bonito e Cachoeiras de Macacu, conforme pode ser visto na Tabela 3.2, ordenada pela população segundo o censo demográfico de 2010 (Fonte: IBGE).

²<http://www.teleco.com.br/num.asp>, visto em agosto de 2017.

Tabela 3.2: Cidades da área de estudo.

Cidade	Área(km ²)	População Censo 2010	Densidade populacional em 2010 (/km ²)	Município pertence à RMRJ?
Rio de Janeiro	1.199,83	6.323.037	5.018	☒
São Gonçalo	247,71	1.013.901	4.014	☒
Nova Iguaçu	519,16	797.212	1.518	☒
Duque de Caxias	467,62	775.456	1.840	☒
Niterói	133,92	487.327	3.769	☒
Belford Roxo	77,82	469.261	5.940	☒
São João de Meriti	35,22	459.356	13.200	☒
Magé	388,5	228.150	590	☒
Itaboraí	430,37	218.090	514	☒
Mesquita	41,48	168.403	4.839	☒
<i>Teresópolis</i>	770,6	163.746	759	☐
Nilópolis	19,39	157.483	8.118	☒
Queimados	75,7	137.938	1.791	☒
Maricá	362,57	127.461	1.379	☒
Itaguaí	274,43	109.091	499	☒
Japeri	81,87	95.391	1.151	☒
Seropédica	283,77	78.183	275	☒
<i>Rio Bonito</i>	456,46	55.551	122	☐
<i>Cachoeiras de Macacu</i>	953,8	54.273	257	☐
Guapimirim	360,77	51.487	143	☒
Paracambi	179,77	47.124	262	☒
<i>Mangaratiba</i>	356,41	36.456	190	☐
Tanguá	145,5	30.731	209	☒
Totais	7.862,643	12.085.108	média=2.452	

Estudos recentes, [46] [65], já mostraram que é possível caracterizar estatisticamente as trajetórias humanas com dados de CDR através de técnicas apropriadas. Os dados do censo foram necessários para se estabelecer o modelo geográfico de forma aderente ao sócio-econômico da população. Os dados do PDTU tornaram

possível a validação da metodologia. O conjunto das bases será pormenorizado a seguir, subdividindo-se em: *a)* dados de CDR, *b)* dados selecionados a partir do censo demográfico do IBGE, e *c)* dados de viagens disponibilizados pela Secretaria Estadual de Transporte do Rio de Janeiro, constantes no Plano Diretor de Transporte de 2013.

3.1 Os dados de CDR

A base de dados de telefonia móvel foi cedida por uma operadora de telefonia através de um projeto via FAPERJ, compreendendo o ano de 2014 e somando 2,1 bilhões de registros referentes a 2,9 milhões de assinantes atendidos por 1.078 torres de telefonia abrangendo toda a área 21. Apenas os dados de chamadas de voz realizadas foram disponibilizados para o estudo, de forma que o conjunto de dados não contém informações adicionais de tráfego como chamadas recebidas e mensagens de texto (SMS). Sobre a localização das chamadas, foram disponibilizados apenas os dados de localização do usuário que teve a iniciativa de chamada, não havendo informação sobre a localidade de quem a recebeu.

As informações de cada registro são mostradas na Tabela 3.3.

Campo	Descrição
Dia	O dia do registro.
Hora	A hora do registro.
Duracao	A duração da chamada.
Ddd_orig	O DDD da estação de origem da chamada.
Num_orig	O identificador ID criptografado da estação origem da chamada.
Ddd_dest	O DDD da estação do destino.
Num_dest	O ID criptografado da estação destino da chamada.
Cell_id_orig	O código da antena que atendeu a chamada na origem.
Tp_trafego	Tipo de tráfego: roaming internacional sms etc.
Hold_orig	Nome da operadora que processou a chamada na origem.
Hold_dest	Nome da operadora que processou a chamada no destino.

Tabela 3.3: Dados de um registro de chamada telefônica (CDR)

Três etapas de pré-processamento ELT³ foram necessárias: *a)* eliminação imediata de 10 bilhões de registros repetidos, *b)* desconsideração de usuários com comportamento anômalo, com mais de 100 chamadas por dia ou menos de 10 chamadas por ano, e *c)* geo-referenciamento de cada chamada, atribuindo-se a posição da torre de telefonia que atendeu a chamada.

O embasamento para a desconsideração do comportamento anômalo *b)* levou em conta a distribuição de probabilidade da quantidade de chamadas realizadas por *ID*, conforme pode ser visto na Figura 3.2(a). A distribuição de probabilidade da quantidade de vizinhos também chamou a atenção, ver Figura 3.2(b), onde alguns *IDs* realizaram chamadas para mais de 10.000 *IDs* distintos. Se supõe que tais *IDs* com tamanha quantidade de vizinhos sejam pertencentes à centrais de atendimento, ou telefones compartilhados. Essas distribuições apresentam comportamento similar ao de outros estudos envolvendo dados de CDR, ver ONNELA *et al.* [66], BLONDEL *et al.* [34] e NEWMAN [67].

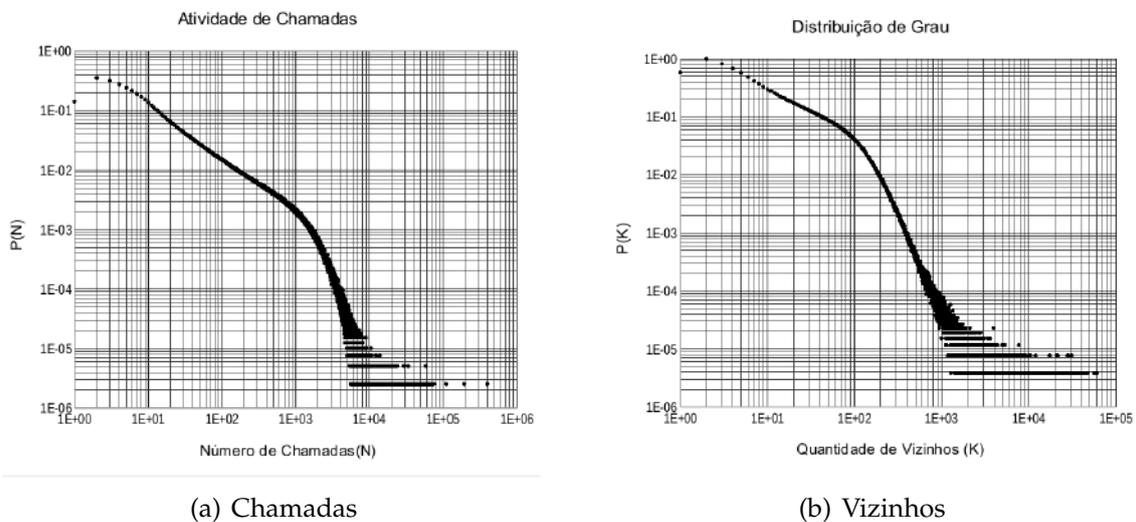


Figura 3.2: Distribuição de probabilidade de chamadas (a) e de vizinhos (b) cobrindo o ano de 2014 da base de CDR. Fonte: Elaboração própria.

As estatísticas básicas dos dados após as três etapas de pré-processamento estão descritas na Tabela 3.4. Observe que, apesar de existirem 1467 torres de telefonia cobrindo a área de estudo, elas não ficam ativas todos os dias do ano, pois

³ELT - Extract Load Transform. Foi necessário carregar todos os registros em seus estados originais para depois realizar o descarte e as transformações.

elas podem entrar em estado de manutenção, conforme procedimentos internos da operadora de telefonia.

A quantidade de identificadores distintos que receberam chamadas durante o ano de 2014 ultrapassa 34 milhões, mais do que toda a população do estado do Rio de Janeiro, o que pode significar chamadas que foram recebidas por identificadores localizados em outros estados ou países. Mais de 13 milhões desses identificadores que receberam chamadas não pertencem à área 21.

Tabela 3.4: Contadores obtidos após primeiro pré-processamento, contemplando todo o ano de 2014.

	Soma Total	Mínimo por dia	Máximo por dia	Média por dia	Desvio padrão
CDRs	2.112.297.808	1.904.078	7.926.462	5.819.002	1.217.536
Ligadores	2.929.094	755.320	1.348.115	1.205.593	109.994
Receptores	34.453.497	1.155.617	2.921.681	2.400.808	374.117
Torres	1467	1052	1078	1067	3,4

A quantidade total de chamadas por dia, e por estação do ano de 2014, pode ser observada na Figura 3.3. Pode-se observar uma redução gradativa na quantidade de chamadas durante o ano, essa redução não necessariamente significa “perda de clientes” da operadora de telefonia, mas simplesmente a tendência das pessoas realizarem menos chamadas telefônicas e mais chamadas através de aplicativos de voz sobre IP, como o caso popular do Whatsapp⁴, por exemplo.

⁴O Whatsapp teve início em 2009, ver <https://pt.wikipedia.org/wiki/WhatsApp>, visto em 3 de abril de 2018.

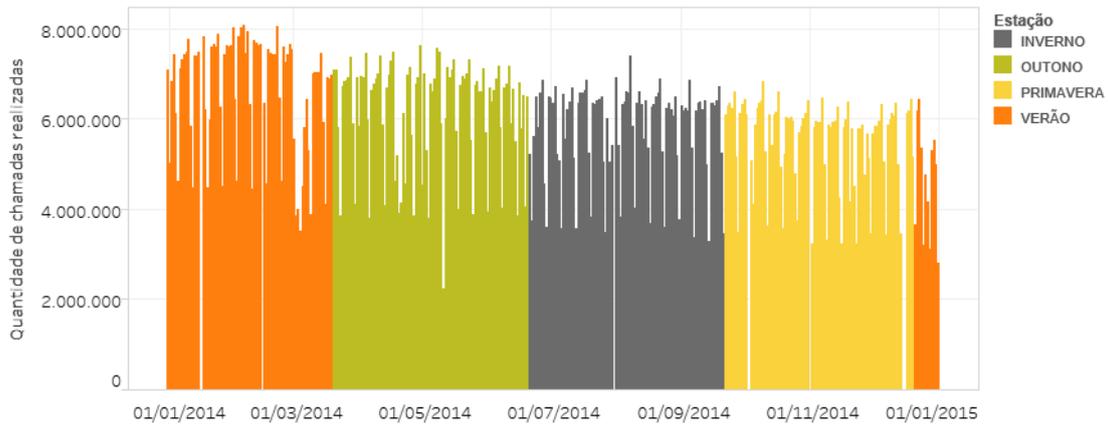


Figura 3.3: Quantidade de chamadas realizadas por dia e por estação do ano.

Em geral, pelo menos 3 antenas são posicionadas a 120° em cada torre, de forma que a cobertura por torre pode ser aproximada por polígonos de Voronoi. A Figura 3.4 mostra a área de estudo no detalhe, e uma visão geral da distribuição espacial das antenas na área de estudo e polígonos de Voronoi correspondentes.

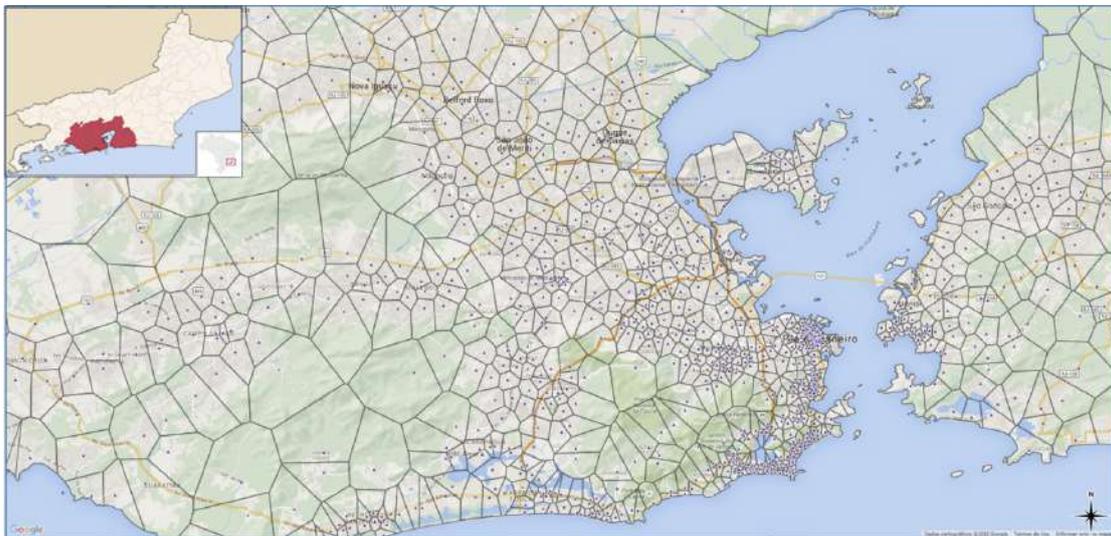


Figura 3.4: Visão geral da área de estudo com foco na distribuição espacial das antenas com seus respectivos polígonos de Voronoi. No detalhe a localização da Região Metropolitana do Rio de Janeiro.

O posicionamento das torres segue objetivos mercadológicos e técnicos e sua distribuição é bastante irregular, como pode ser observado na Figura 3.4. A distribuição espacial das antenas é mais densa nas áreas de maior renda e oferta de empregos. Nas áreas Norte e Oeste, de menor densidade demográfica, a distri-

buição das antenas é menos densa, assim como nos demais municípios da Região Metropolitana, exceto o centro da cidade de Niterói. A baixa densidade de torres de telefonia observada em algumas áreas no meio da zona urbana se justifica devido às montanhas. A geografia da cidade do Rio de Janeiro causa um grande impacto nas rotas de mobilidade da cidade.

A irregularidade da distribuição espacial das antenas em relação a área e a população das cidades também pode ser vista na Tabela 3.5, ordenada pela quantidade de torres de telefonia por km². Também são mostrados valores médios da população coberta por cada torre. O Rio de Janeiro, por exemplo, com uma população de 6.323.037 pessoas é coberto por 707 torres, ou seja, 8.943 pessoas/torre em média. Dentre as cidades que mais ligam, a capital também se destaca com uma média maior que 3 milhões de chamadas por dia.

Tabela 3.5: Estatísticas das torres de telefonia nas cidades contempladas na área de estudo.

Cidade	Torres de telefonia por km ²	População por torre	Média diária de chamadas por cidade
CACHOEIRAS DE MACACU	0,0084	9032	5152
TANGUÁ	0,0109	15241	4175
RIO BONITO	0,0159	7857	8172
JAPERI	0,0164	31084	13422
TERESÓPOLIS	0,0176	9609	33647
ITAGUAÍ	0,0202	15562	41373
GUAPIMIRIM	0,0224	7343	42601
SEROPÉDICA	0,0225	15499	32473
PARACAMBI	0,0261	15337	17192
ITABORAÍ	0,0274	16738	82181
MARICÁ	0,0288	10609	51004
MANGARATIBA	0,0324	3037	29729
MAGÉ	0,0420	14138	284191
NOVA IGUAÇU	0,0596	26513	234355
QUEIMADOS	0,0790	19680	45931
DUQUE DE CAXIAS	0,0957	20831	331472
BELFORD ROXO	0,1382	39075	85699
MESQUITA	0,1806	42073	38677
SÃO GONÇALO	0,2279	16914	607097
NILÓPOLIS	0,3107	26216	74874
NITERÓI	0,4907	6387	351589
SÃO JOÃO DE MERITI	0,5009	25528	133828
RIO DE JANEIRO	0,5604	8943	3387408

Enfim, os dados de CDR constituem uma fonte rica para estudos de mobilidade, contudo precisam ser contextualizados e compreendidos como amostras

da população, para que se mantenha o senso de proporção entre os objetivos mercadológicos das operadoras e o censo demográfico.

3.2 Os dados do censo

O IBGE é a instituição oficial do país para estudos do sócio-econômico da população. A cada 10 anos é realizado um censo nacional com amostras populacionais. São consideradas cerca de 200 variáveis categorizadas por atributos de pessoas, domicílios, migração e mortalidade. Algumas variáveis são novas, como foi o caso da pergunta se a pessoa possui ou não telefone celular, perguntada em 2010; e outras são antigas e comparáveis historicamente, como por exemplo se o entrevistado sabe ler e escrever. Assim como o IBGE procede no Brasil, existem instituições que fazem o mesmo em diversos países, possibilitando comparações internacionais dos mesmos atributos, para maiores informações, consultar a *IPUMS International*⁵.

Os dados do censo do IBGE são hierarquizados de forma que a menor granularidade representativa é a de uma família, que possui um peso amostral de acordo com o método estatístico explicado em BANKIER [68]. Um setor censitário é um conjunto de aproximadamente 200 famílias, independente da área geográfica de abrangência. É possível que um grande condomínio de prédios confinados numa pequena área, assim como um conjunto de bairros em que existam apenas casas, sejam considerados na mesma proporção de um setor censitário; obviamente sob densidades demográficas diferentes.

A hierarquia superior ao setor censitário é o subdistrito, depois o distrito, depois a cidade, depois a mesorregião, depois o estado, e finalmente o país. Dada a amplitude da RMRJ, e conseqüentemente da área 21, o nível de cidade seria o mais adequado para a evolução da pesquisa, contudo os indicadores populacionais (Tab. 3.2) e de média diária de chamadas telefônicas (Tab. 3.5) apontam a cidade do Rio de Janeiro com indicadores muito distantes das demais. Essa dis-

⁵<https://international.ipums.org/international/> visto em maio de 2017.

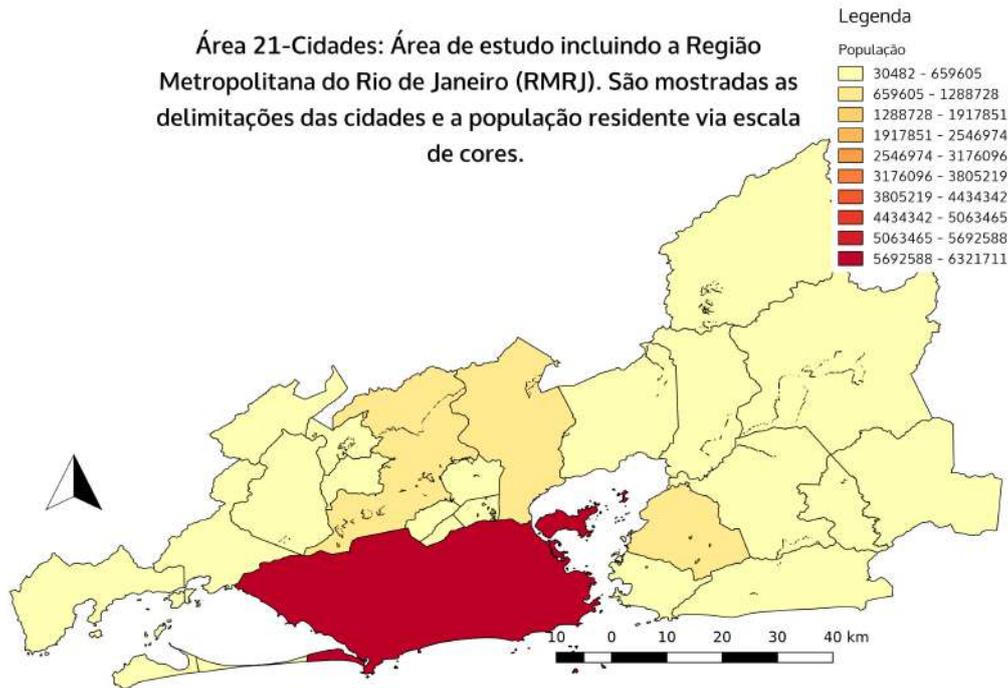


Figura 3.5: Cidades da área de estudo. Fonte: Elaboração própria a partir do censo de 2010.

tinção fica ainda mais clara ao se observar a população projetada no mapa, ver Fig. 3.5.

Uma alternativa, para tornar o particionamento geográfico da área de estudo mais homogêneo, é o uso de um sub-particionamento apenas para a cidade do Rio de Janeiro, por exemplo, subdistritos, ver Fig. 3.6.

O particionamento “misto”, ou seja, as cidades da área 21 e a cidade do Rio de Janeiro sob o nível de subdistritos (*conforme visto na Figura 3.6*), tende a se adequar melhor ao universo proposto de estudo, pois torna-se possível observar os indicadores fornecidos pelo censo sob um nível de precisão geo-espacial maior dentro da capital, como por exemplo a renda per capita. Segundo o IBGE, a renda domiciliar per capita é calculada através da razão entre o total de rendimentos e o total de moradores da residência, esse indicador pode ser visto, em R\$, na Figura 3.7.

O particionamento geográfico, visto nas Figuras 3.6 e 3.7, tende a tornar-se um modelo para toda a pesquisa, fazendo com que todas as operações geo-espaciais

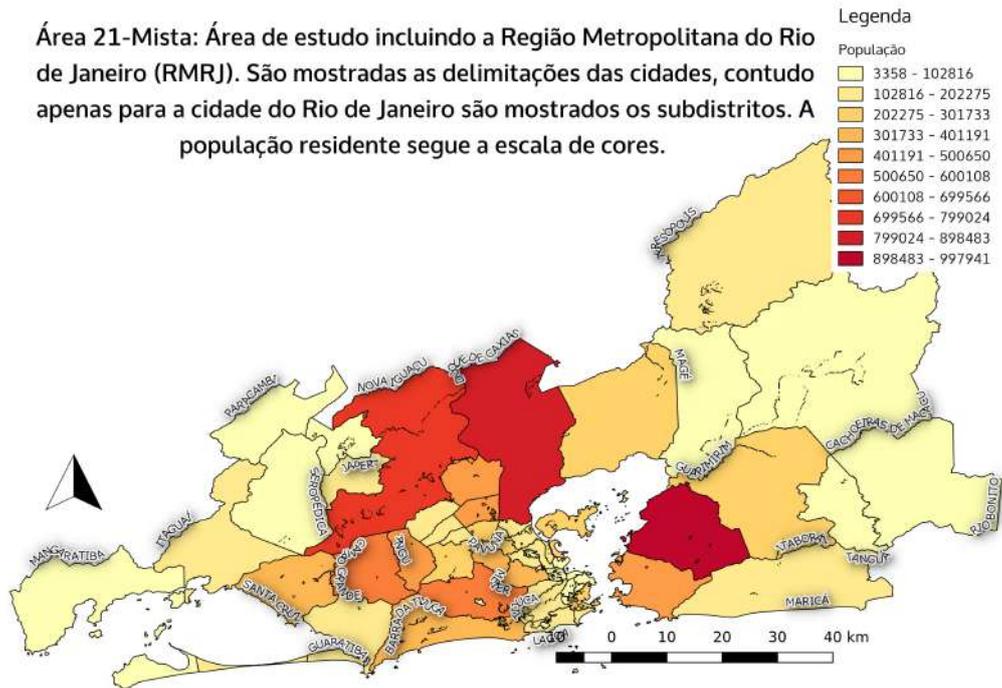


Figura 3.6: Cidades da área de estudo, sendo a cidade do Rio de Janeiro particionada em subdistritos. Fonte: Elaboração própria a partir do censo de 2010.

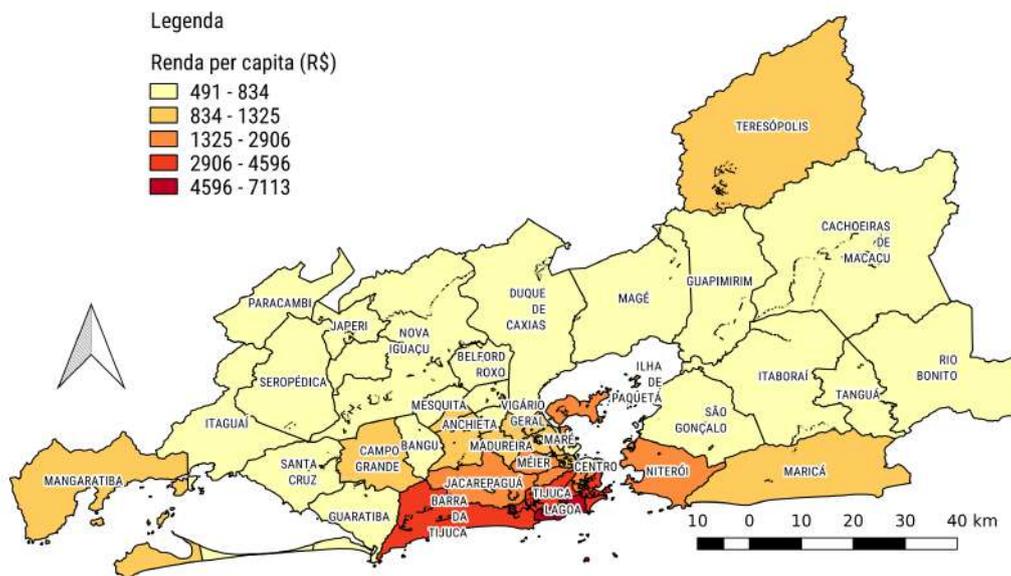


Figura 3.7: Renda domiciliar per capita das cidades da área 21, apenas a cidade do Rio de Janeiro está particionada em subdistritos. Fonte dos dados: censo de 2010.

inerentes ao estudo sigam esse modelo. Um discussão sobre esse modelo será apresentada no final deste capítulo.

Enfim, as informações disponibilizadas pelo IBGE consistem em uma grande fonte pública de dados sócio-econômicos da população.

3.3 Os dados do PDTU

Conforme visto na revisão bibliográfica (Seção 2.1.2), o PDTU é o nome adotado em algumas regiões metropolitanas para definir um plano de transporte, geralmente disponibilizado pela secretaria de transportes responsável. No caso da RMRJ, os dados do PDTU de 2013 [4] foram usados com a finalidade de validar a quantidade de viagens inferida pela metodologia do presente estudo. As discussões sobre a validação serão apresentadas no capítulo de resultados.

Em estudos que possuem atributos geoespaciais, como é o caso do PDTU, é comum o estabelecimento de um particionamento geográfico padronizado; isso possibilita a realização de cruzamentos e agregações de atributos. Essa padronização pode adequar-se a fontes externas, como é o caso do PDTU. O PDTU de 2013 adotou um modelo geográfico a partir de composições de subdistritos do IBGE, chamado macrozona de tráfego.

Macrozonas de tráfego

O PDTU da RMRJ é a referência oficial para estudos da área de transportes, feito pela Secretaria de Transportes do Estado do Rio de Janeiro. Considerando os relatórios de 2003 [3] e de 2013 [4] é possível observar uma mudança na granularidade de estudo. Anteriormente os polígonos não respeitavam inteiramente as delimitações de subdistritos/cidades do censo, tornando árdua a tarefa de relacionar todas as macrozonas de tráfego à unidades do IBGE. Já no último relatório as partições mudaram, de tal forma que uma macrozona ou é um conjunto de subdistritos, ou um conjunto de cidades. Essa mudança tornou o particionamento geográfico do PDTU mais compatível com o do IBGE, facilitando o georeferenciamento aos dados sócio-econômicos.

Sendo a macrozona um particionamento geográfico planejado para ser aderente aos dados do censo, observe como o atributo *renda per capita em salários mínimos* foi inserido no contexto das principais vias de acesso da RMRJ na Figura 3.8. As maiores rendas foram registradas na região central e oceânica de Niterói e a na Zona Sul da cidade do Rio de Janeiro.

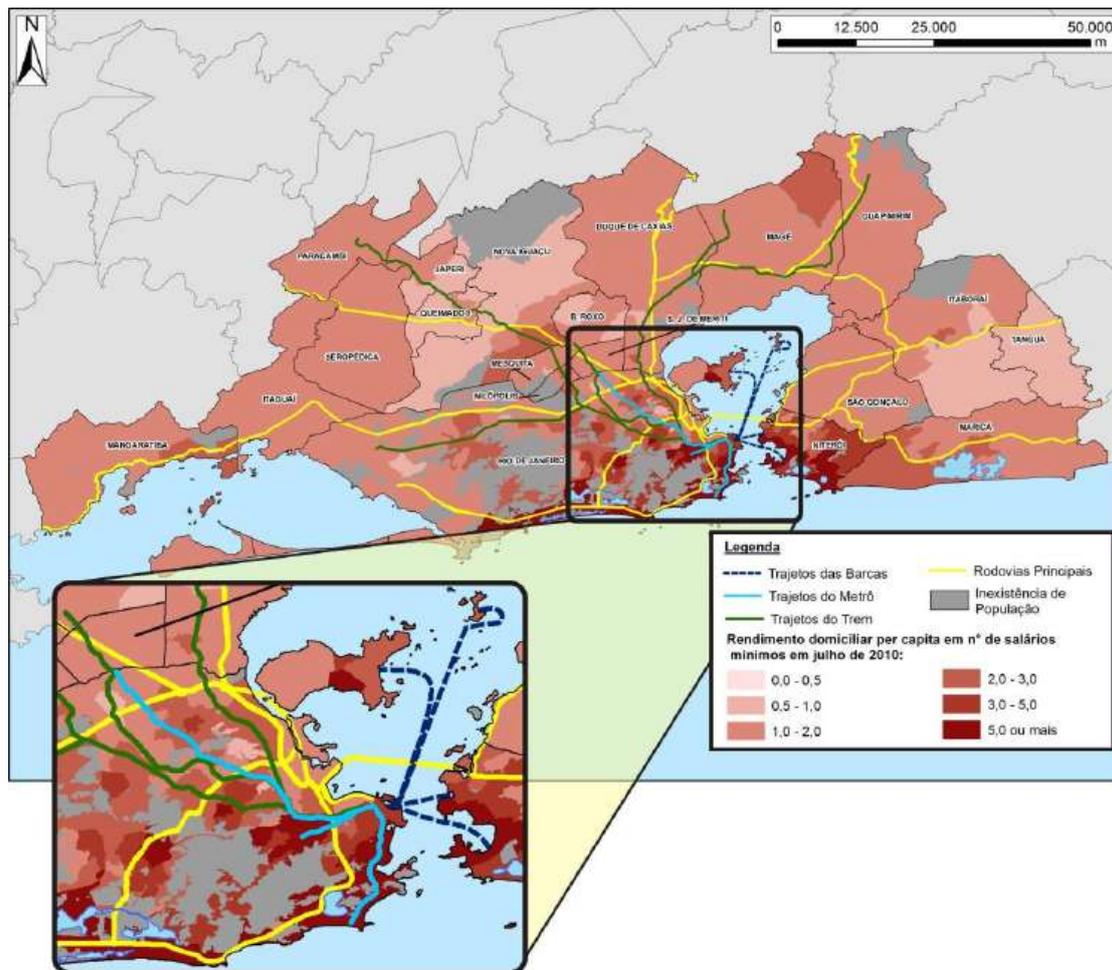


Figura 3.8: Área incluída na pesquisa do PDTU de 2013 com destaque para os maiores corredores de transporte. Figura extraída da página 65 do PDTU 2013 [4].

A cidade do Rio de Janeiro é responsável por originar 61,3% das viagens na RMRJ, seguida de Nova Iguaçu com 6,4%. Mais da metade das viagens, 58,8%, são feitas da seguinte forma: 29,4% a pé e 29,5% de ônibus municipal, seguido por 11,2% de automóvel (como condutor) e 5,4% como passageiro .

As macrozonas de tráfego visam estudar a mobilidade da população a partir dos maiores corredores viários. A Figura 3.9 mostra a delimitação de polígonos feita segundo o Plano Diretor de Transporte Urbano (PDTU) da RMRJ. Observe que as áreas ficaram sob uma agregação maior se comparadas com as unidades do particionamento misto (Fig. 3.6). A cidade do Rio de Janeiro foi particionada em apenas quatro macrozonas, ao mesmo tempo que foi subdividida em 32 sub-distritos segundo o censo demográfico.

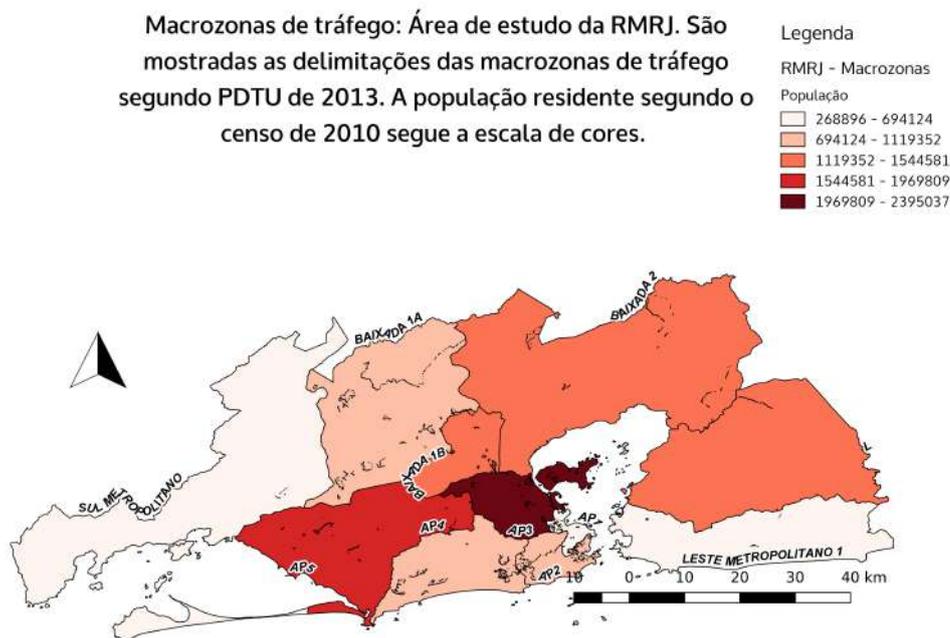


Figura 3.9: Macrozonas de tráfego da RMRJ segundo o PDTU de 2013.

A relação entre as unidades geográficas desta pesquisa (Fig. 3.6) e as macrozonas de tráfego (Fig. 3.9) podem ser vistas na Tabela A.1 (Apêndice A.1). A área 21 é uma região maior do que a RMRJ, as regiões excedentes foram mantidas propositalmente no estudo, pois é comum que essas delimitações mudem devido a adequações de zoneamento, como foi o caso do PDTU de 2013 [4] em relação ao de 2003 [3]. A título de padronização, as regiões não pertencentes à RMRJ estão denominadas por "EXTRA URBS".

Enfim, considerando que os microdados do PDTU são disponibilizados em nível de zonas de tráfego, e que pode-se agregá-los ao nível de subdistrito, e conseqüentemente de cidade, é possível seguir com uma definição própria de modelo geográfico sem perder a capacidade de validação. Será mostrada, a seguir, uma discussão sobre o particionamento “misto”, que será a base do modelo geográfico da metodologia.

Capítulo 4

Estimativas de matrizes de origem-destino

“Relate suas mágoas e seus desejos, seus pensamentos passageiros, sua fé em qualquer beleza - relate tudo isto com íntima e humilde sinceridade.”

— Rainer Maria Rilke, Cartas a um jovem poeta.

Conteúdo

4.1	Padronização geográfica	65
4.2	Particionamento espacial	68
4.3	Domicílio presumido dos usuários de telefonia	69
4.3.1	Fatores de expansão populacional	75
4.4	Matrizes de origem-destino quotidianas	80
4.4.1	Matriz O-D NxN (<i>de todos locais para todos locais</i>)	81
4.4.2	Matrizes O-D HBO e NHB	86
4.5	Matriz de origem-destino extraordinária	88

Como visto na revisão bibliográfica, diversas pesquisas envolvendo dados de CDR já abordaram o tema de estimativa de matrizes de origem-destino (O-D). Sejam nos estudos tradicionais da área de transporte, ou sejam nos que envol-

vem dados de telefonia, fica clara a necessidade do estabelecimento de áreas que delimitem o que é uma origem e o que é um destino.

Cada região pode ser estudada a partir de uma finalidade própria, ligada às características de interesse do estudo. O censo demográfico irá tratar de atributos sócio-econômicos da população, por isso faz sentido que as delimitações de interesse agrupem populações de características sócio-econômicas homogêneas. Da mesma forma, a área de transportes tem interesse na mobilidade da população, por isso faz sentido que as delimitações correspondam à zonas de tráfego que levem em conta a malha viária. A área de segurança pública irá considerar a área de abrangência das delegacias, e assim por diante.

O dado de CDR é vinculado às torres de telefonia móvel, que não seguem padrões amostrais simples, e conforme visto na revisão bibliográfica (Seção 2.2, Figuras 2.5 e 2.6), possuem uma distribuição geográfica irregular.

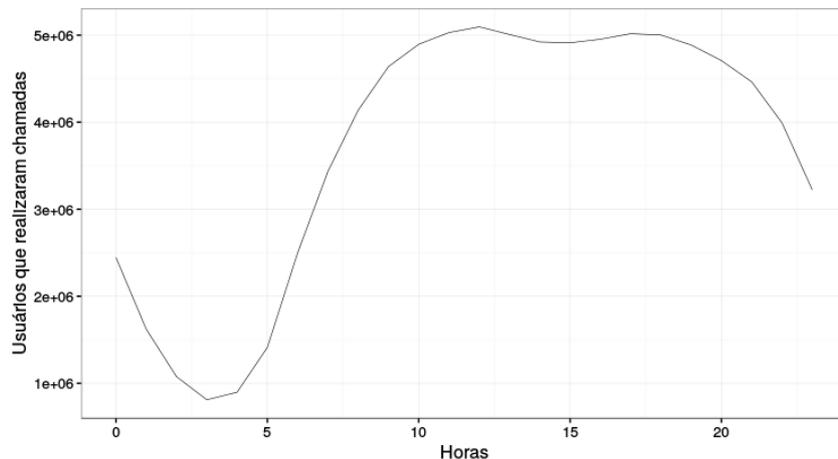


Figura 4.1: Quantidade de usuários ativos por hora (média de todos os dias).

A distribuição temporal pode variar significativamente durante as horas do dia (Fig. 4.1), assim como a carga de trabalho de cada torre (Fig. 4.2). Existem antenas que atenderam apenas uma chamada e outras que atenderam mais de 15 milhões.

Por isso faz-se necessária a definição de um particionamento espaço-temporal, a ser adotado no curso de toda a pesquisa. A seguir será mostrado esse processo de particionamento.

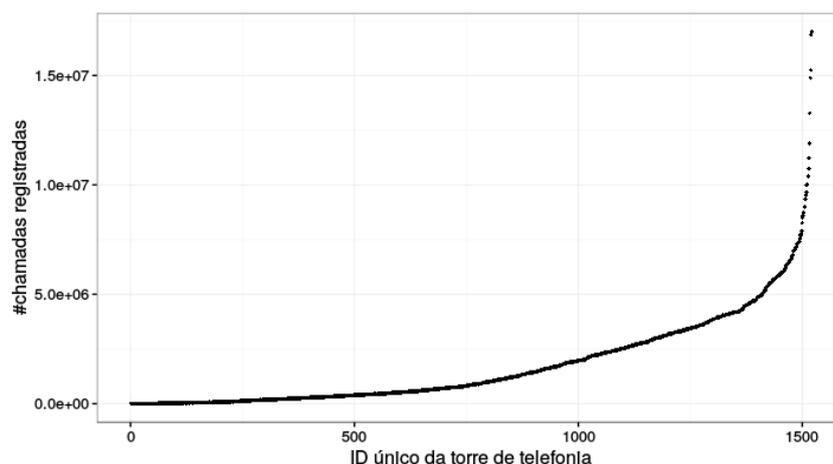


Figura 4.2: Quantidade média de chamadas processadas por torre de telefonia.

4.1 Padronização geográfica

A fluidez e a aplicabilidade da pesquisa está ligada a uma padronização geográfica. Essa definição tende a tornar-se parte inerente do estudo e ser usada em diversas etapas de processamento, de forma que a metodologia tende a ser *regida* por essa padronização. Um exemplo de padronização geográfica feita a partir de dados de telefonia móvel pode ser vista em LOUAIL *et al.* [69].

A padronização geográfica precisa levar em conta as posições das torres de telefonia móvel, pois os dados são todos georeferenciados a elas. Conforme visto na revisão bibliográfica (Seção 2.2), cada polígono de voronoi corresponde a área **aproximada** de cobertura das antenas. A posição geográfica real dos indivíduos está ligada diretamente à das antenas. Quanto menor a densidade de antenas, menor a precisão de onde as pessoas realmente estão. A forma encontrada para dissolver essas limitações foi agrupar torres de telefonia com a intenção de respeitar os limites entre os subdistritos e as cidades, os limites do particionamento geográfico “misto”.

Observe, na Figura 4.3, como fica a disposição dos polígonos de voronoi sobre alguns subdistritos da cidade do Rio de Janeiro. Nas regiões em que a densidade de antenas é maior, torna-se mais fácil a intenção de respeitar os limites entre os subdistritos, no entanto a aproximação em regiões de baixa densidade tende a ser

mais incerta, de forma até coerente com a incerteza de cobertura em regiões tão vastas.

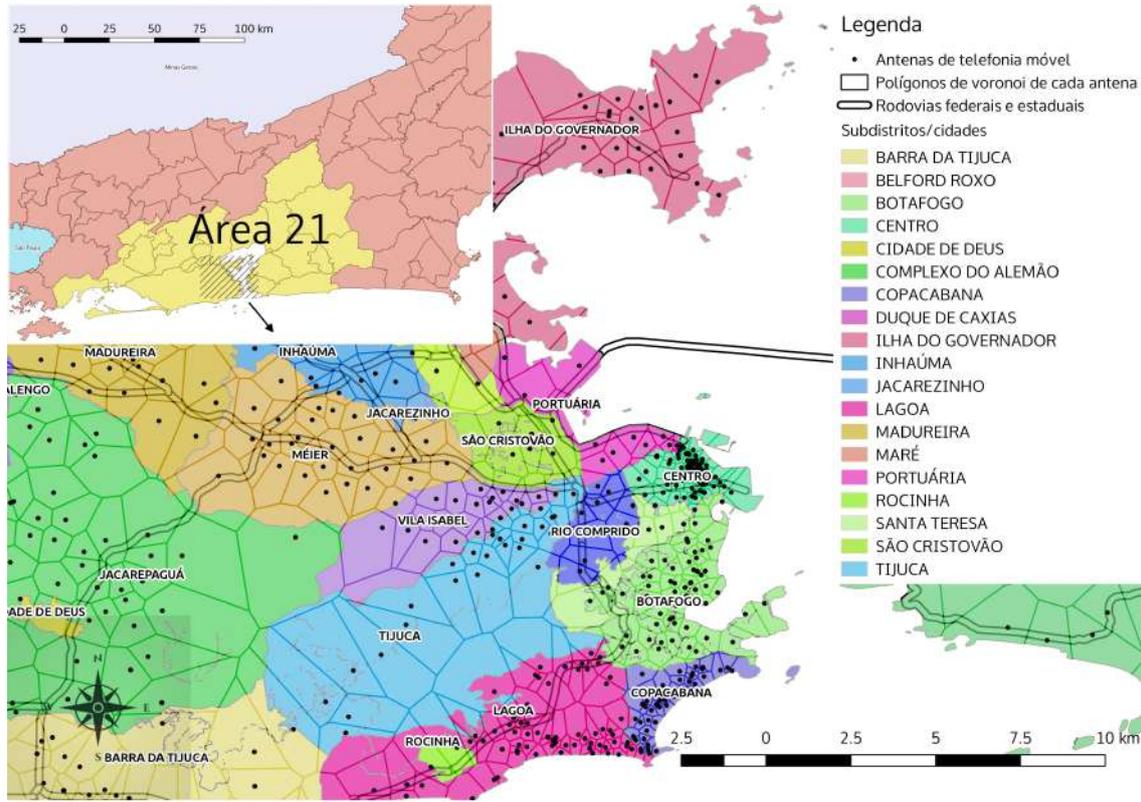


Figura 4.3: Recorte da área de estudo mostrando as antenas sobre alguns subdistritos.

As delimitações estabelecidas a partir do conjunto de polígonos de voronoi, na intenção de respeitar os limites estabelecidos pelo IBGE, formam uma nova *visão* da área de estudo. A Figura 4.4 mostra como fica essa nova *visão*, destacando a densidade de torres de telefonia por Km². O centro do Rio de Janeiro é a região com maior densidade de torres de telefonia, assim como as regiões no entorno.

A região central pode ser vista de forma ampliada na Figura 4.5, destacando e quantidade média de chamadas por dia, e por região, realizadas durante um ano. Dentre as unidades geográficas pré-definidas para o estudo, a cidade de São Gonçalo possui a maior média diária de realização de chamadas.

Enfim, foram mostradas as diversas bases de dados usadas na pesquisa, culminando na discussão sobre o estabelecimento de um particionamento geográfico próprio, que seja capaz de “ligar” todas elas. Dada a importância dessa definição,

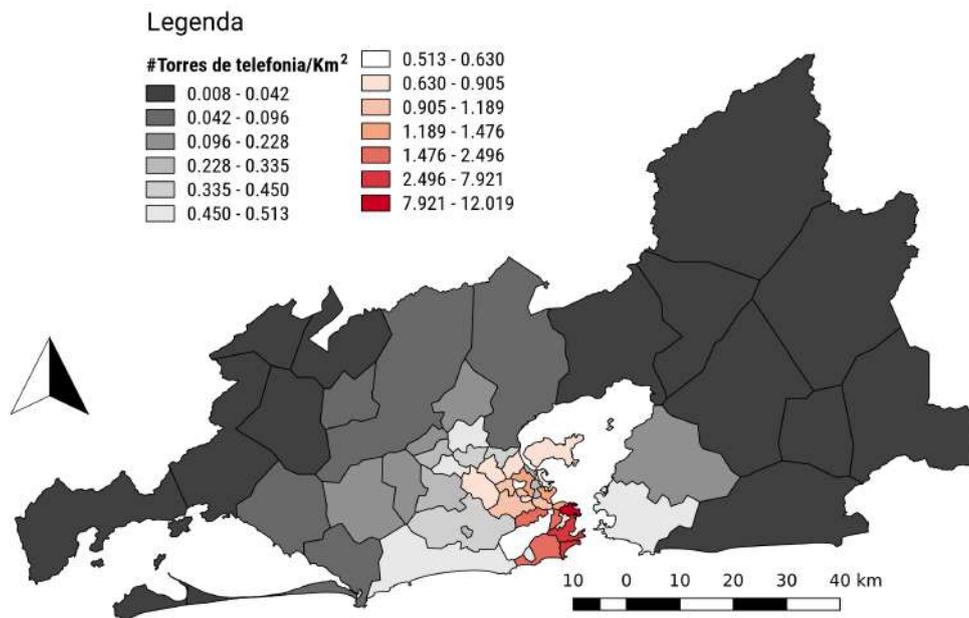


Figura 4.4: Quantidade de torres de telefonia por Km², por unidade geográfica do particionamento misto.

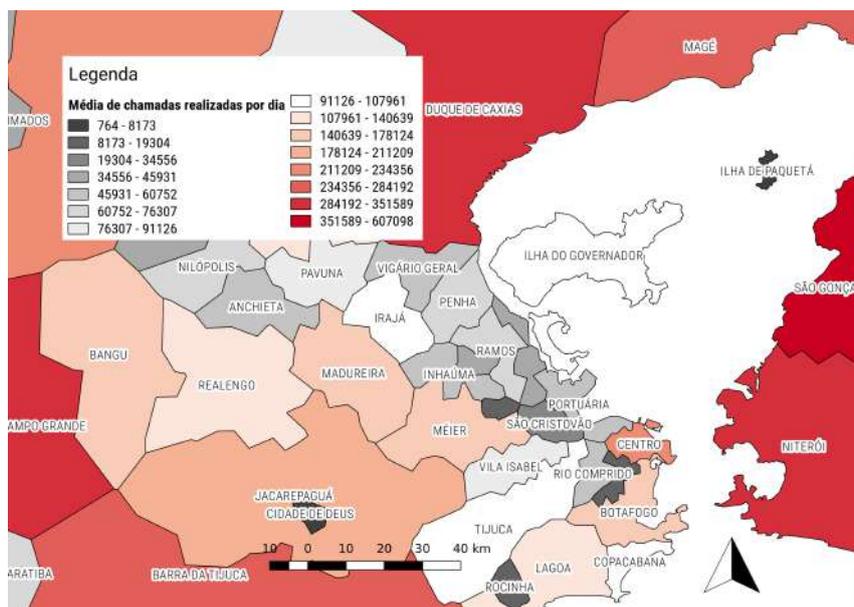


Figura 4.5: Quantidade média de chamadas por dia por região.

e da forma como ela está *entrelaçada* em toda a pesquisa, a abordagem sobre o particionamento irá continuar no próximo capítulo sob o um enfoque mais metodológico.

4.2 Particionamento espacial

Em muitos estudos a partir de dados de CDR, é adotado um particionamento baseado em *grid* regular, conforme visto na Seção 2.2, artigos [7], [70], [47] e [51] (Figura 2.7). A solução é simples e de fácil generalização, contudo é de difícil interpretação e requer um ajuste de geometria para ser relacionado com informações demográficas e outras fontes de dados.

Este trabalho propõe um particionamento espacial que considera as características sócio-econômicas da população. A ideia é usar as divisões territoriais já adotadas pelo IBGE por conta do censo demográfico. O particionamento do IBGE segue agregações de setores censitários, que formam subdistritos, distritos, cidades e estados. O IBGE adota padrões internacionais para esse tipo de particionamento, de forma que a metodologia é reproduzível em outras cidades e regiões metropolitanas (Fig. 4.6).



Figura 4.6: Particionamento espacial.

A Figura 4.7 ilustra o particionamento espacial adotado no estudo, focando nos locais de maior concentração de registros e de maior densidade populacional. O estudo foi conduzido sob o nível de subdistritos para a cidade do Rio de Janeiro e cidade para os demais locais da RMRJ, resultando em 55 unidades geográficas. Esse particionamento espacial simplifica a visão da RMRJ para o planejamento de transporte. Cada unidade geográfica representa características sócio-econômicas homogêneas, respeitando as próprias definições do IBGE.

A RMRJ tem uma geografia peculiar, com morros ao redor da baía da Guanabara, tal que algumas áreas geograficamente próximas podem não possuir conectores, ou serem conectadas por túneis e pontes. Por esse motivo o cálculo das distâncias entre as partições de estudo precisam considerar essas características.



Figura 4.7: Visão do particionamento espacial, com ênfase nos locais em houve maior quantidade de registros.

Uma vez definido o particionamento espacial, podem-se inferir origens e destinos, contudo os usuários de telefonia precisam pertencer a um determinado local de residência, principalmente para poder realizar comparações e validações com outras bases de dados. Do contrário, o estudo seria restrito apenas aos usuários de uma operadora de telefonia.

4.3 Domicílio presumido dos usuários de telefonia

O cálculo de domicílio presumido na metodologia é posterior ao particionamento espaço-temporal, e anterior a elaboração das matrizes de origem-destino, ver Figura 4.8. Os usuários de telefonia precisam ser multiplicados por um fator de expansão, para que se aproximem da “população” itinerante. Coincidentemente, o fator de expansão pode ser compreendido como o “*market share*” da operadora de telefonia.

A unidade geográfica mais visitada entre 19h e 6h do dia seguinte em dias úteis, ou o dia inteiro em domingos e feriados, foi considerada como o domicílio presumido de cada usuário. Também foi necessário que o local visitado com



Figura 4.8: Posicionamento metodológico do domicílio presumido.

maior frequência pelo usuário, correspondesse no mínimo a 50% das visitas em relação a todos os locais visitados.

De forma a detalhar as explicações, o processo de cálculo de domicílio presumido foi dividido em etapas (Fig. 4.9), obedecendo uma sequência lógica que chega ao resultado no último passo. Essa forma de apresentar a metodologia será seguida até o final do capítulo. Observe que, a maioria das etapas foi resolvida através da linguagem SQL ANSI, e sendo assim, partes de código dessa linguagem serão expostas na intenção de tornar o entendimento mais preciso.

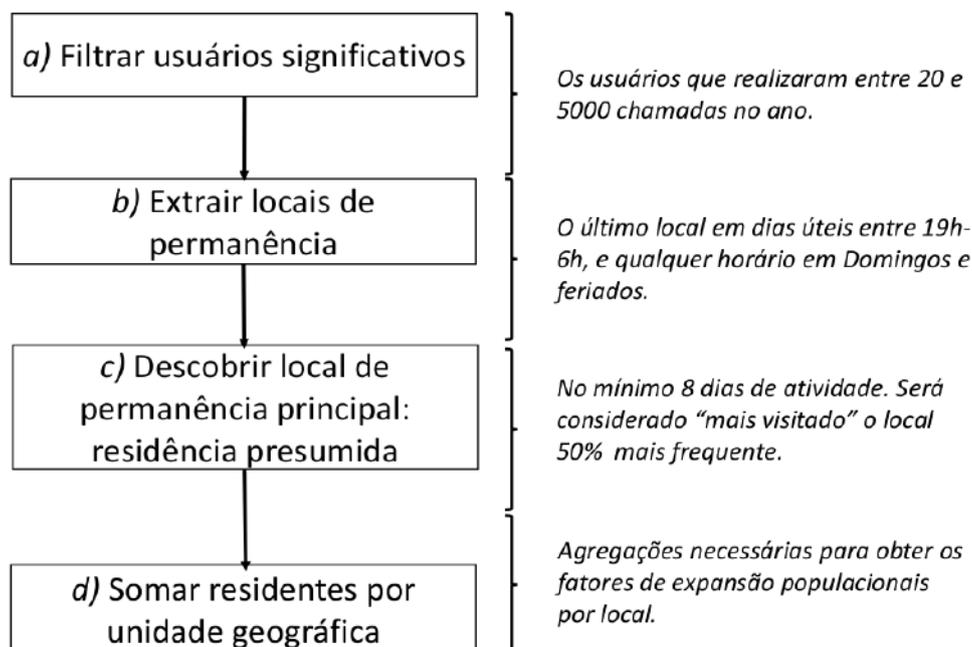


Figura 4.9: Fluxo de passos para cálculo do domicílio presumido, desde o dado "bruto" de CDR até a quantidade de usuários residentes por unidade geográfica.

A inferência do domicílio dos usuários de telefonia em si (Fig. 4.9), ocorre basicamente em três etapas (*a, b e c*), sendo a quarta etapa (*d*) apenas uma “finalização”, de onde são obtidos os valores que darão suporte aos fatores de expansão. Cada um desses itens será detalhado a seguir:

a) Usuários significativos: tal como o posicionamento geográfico dos usuários está ligado a realização de chamadas telefônicas, um usuário é significativo quando possui uma quantidade mínima de 20 chamadas realizadas em um ano, assim como uma quantidade máxima de 5000. Esses delimitadores foram baseados nas próprias distribuições (ver Fig. 3.2). Valores fora desse intervalo são insuficientes para inferir um local de permanência, ou são considerados “anormais”, tais como aparelhos compartilhados ou centrais telefônicas.

SQL 4.3.a

```
_____ Filtro de usuários significativos _____  
1 create table ligadores_significativos as  
2 select ID  
3 from (  
4   select count(*) as count, ID  
5   from CDR  
6   group by ID) as s0  
7 where count > 20 and count < 5000;
```

b) Locais de permanência: a extração de locais de permanência visa compreender um comportamento típico de mobilidade individual, considerando que a maioria das pessoas sai de casa em dias úteis e retorna para dormir, e tende a permanecer em casa nos Domingos e feriados. Assim sendo, foram considerados locais de permanência aqueles em que a pessoa ligou entre 19h-6h em dias úteis e qualquer horário em Domingos e feriados.

Para melhor elucidar esta etapa, a Figura 4.10 mostra a quantidade de locais de permanência *versus* a quantidade de usuários (em escala logarítmica).

Observa-se que na maioria dos casos os usuários são detectados em apenas um local de permanência (parte superior do gráfico), considerando os filtros pré-estabelecidos. Observe também que nem todos os usuários contados nesse gráfico tiveram o local de residência detectado.

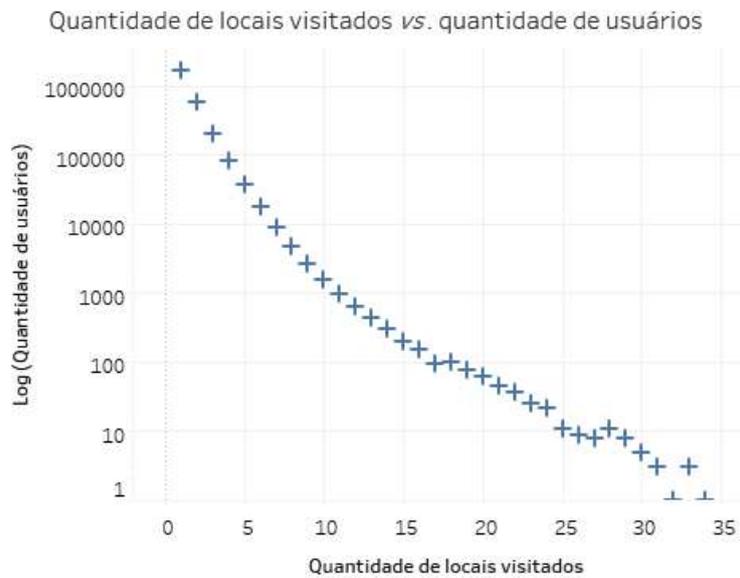


Figura 4.10: Quantidade de locais de permanência detectados sob os filtros estabelecidos, segundo quantidade de usuários (em escala logarítmica).

A seguir, (SQL 4.3.b), será mostrada a instrução responsável por esta etapa. Observe que o local de permanência já está subordinado ao particionamento geográfico anteriormente definido, representado pelo atributo `PARTICAO_GEOGRAFIA`, abaixo.

Extração de local de permanência

```
1 SELECT ID, ..., PARTICAO_GEOGRAFIA
2 FROM CDR ...
3 WHERE ... DIA IN (SELECT dia
4                   FROM feriado)
5 OR EXTRACT (dow FROM DIA) = "DOMINGO"
6 OR ( EXTRACT (HOUR FROM HORA) > 19
7 OR EXTRACT (HOUR FROM HORA) < 6)....
```

c) Residência presumida: dentre os locais de permanência descobertos na etapa anterior, apenas um será considerado como local de residência. O primeiro critério é que o usuário tenha realizado chamadas em no mínimo 8 dias, independente do local. O segundo critério é que o local mais frequente corresponda a no mínimo 50% entre todos os locais de permanência encontrados.

Para elucidar melhor esta etapa, a Figura 4.11, mostra a diferença entre a quantidade de visitas em que os usuários foram detectados em qualquer local e o local inferido como residência. No topo, se vê que aproximadamente 400 mil usuários foram detectados em apenas um local, ou seja, *total de visitas em locais - visitas em local de residência = 0* . Na cauda, um usuário chegou a diferença de 415 visitas. Esse gráfico mostra a contagem apenas dos usuários que tiveram o local de residência detectado.

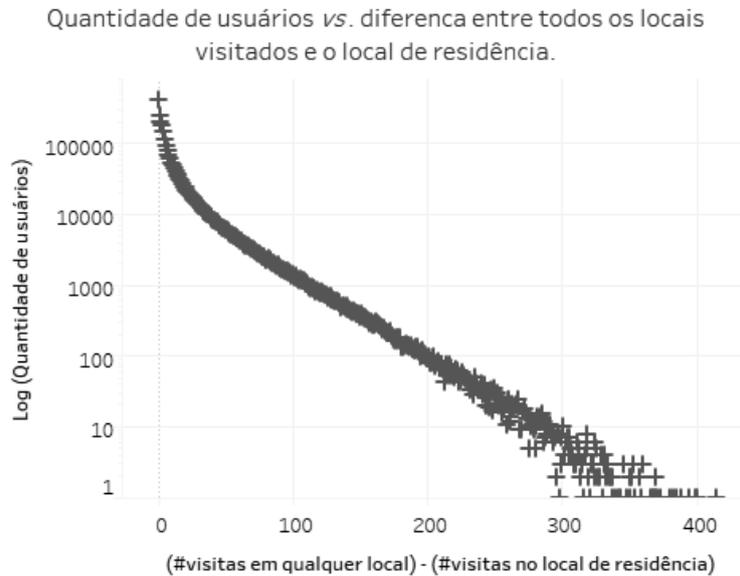


Figura 4.11: Diferença entre a quantidade total de detecções e quantidade de detecções no local inferido como residência.

O pré-requisito de contar a quantidade de dias em que o usuário foi detectado em algum local, é feito através da instrução `contar_dias_em_locais`, abaixo (SQL 4.3.c1).

SQL 4.3.c1 _____ Versão completa em A.2.2.

Total de dias em locais

```

1 SELECT ID, SUM (n_dias) AS total_dias_em_locais
2 FROM ( .... FROM ultima_posicao_dias...
3     GROUP BY ID, PARTICAO_GEOGRAFIA)...
4 AS ultima_posicao...

```

O cálculo que finaliza o local de residência dos usuários de telefonia é feito através da instrução abaixo (SQL 4.3.c2). Observe que as duas condições principais estão nas duas últimas linhas: `total_dias_em_locais >= 8`, para assegurar que o usuário tenha sido detectado em 8 dias ou mais, e `dias_local_mais_visitado > total_dias_em_locais / 2`, para que a quantidade de dias no local mais visitado seja maior do que a metade de todos os dias de detecção. Ou seja, mais de 50% das vezes.

SQL 4.3.c2 _____ Versão completa em A.2.3.

```
_____ Residência presumida dos usuários _____  
1 ...select ID, PARTICAO_GEOGRAFIA, dias_local_mais_visitado, total_dias_em_locais  
2 FROM ( ...  
3 SELECT ... FROM ... ) ....  
4 where total_dias_em_locais >= 8  
5 and dias_local_mais_visitado > total_dias_em_locais / 2;
```

d) Residentes por unidade geográfica: trata-se do resultado final do cálculo do domicílio presumido, agregando os usuários detectados pelas respectivas unidades geográficas. Essa agregação é obtida através da instrução abaixo (SQL 4.3.d).

SQL 4.3.d _____

```
_____ Total de usuários por local _____  
1 create table groupby_local as  
2 select PARTICAO_GEOGRAFIA, count(*) as pop_usuarios  
3 from residencia_presumida  
4 group by PARTICAO_GEOGRAFIA;
```

O resultado do algoritmo de domicilio presumido é, para cada usuário na base, um local de residência pertinente a uma unidade geográfica definida no particionamento espacial da área de estudo. Os usuários que não tiveram seus domicílios presumidos identificados serão ignorados no cálculo de todos os processamentos posteriores. De 2.929.094 usuários, 2.593.911 tiveram o seu local de residência descoberto através da metodologia demonstrada.

4.3.1 Fatores de expansão populacional

De forma geral, os fatores de expansão populacional servem à finalidade de expandir uniformemente o “*market share*” da operadora de telefonia na área de estudo. Ou seja, aproximar a quantidade de pessoas residentes em um região,

a partir da quantidade de clientes da operadora detectados como residentes na mesma região, isso poderia ser compreendido como “*market share*” local. O *market share* “global” seria o valor médio entre todos os locais.

Após o particionamento espaço-temporal, as partições geográficas segundo o IBGE, e a residência presumida, é possível estabelecer fatores de expansão à população. Esses fatores (ver Figura 4.12) podem ser fixos ou adaptativos; o primeiro considera um mesmo fator k_l para o local l para qualquer dia na base de dados, o último leva em conta um valor “adaptado” $k(OD, d)$ para cada par OD no dia d .

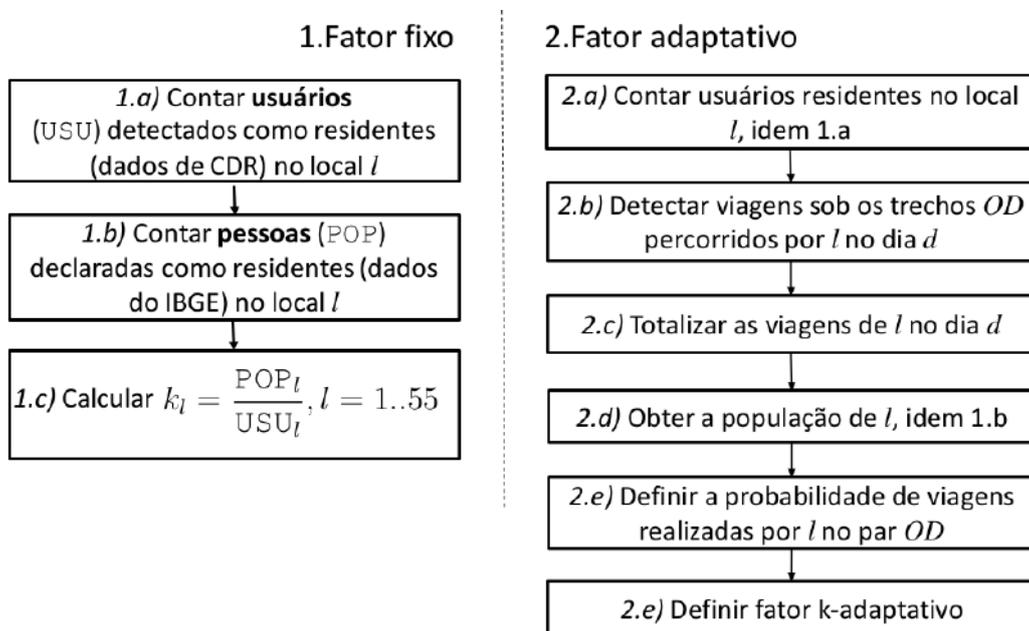


Figura 4.12: Fatores de expansão populacional.

A seguir será apresentada uma descrição das etapas mostradas na Figura 4.12.

4.3.1.1 Fator de expansão fixo k_l

4.3.1.1.a Apenas usuários com domicílio presumido: se a unidade geográfica l de residência de um usuário não foi detectada, não há como atribuí-lo a um peso populacional. Assim sendo, são consideradas apenas as chamadas realizadas por usuários com o local de residência inferido.

4.3.1.1.b Contagem populacional segundo IBGE: uma vez que as partições geográficas dos dados de CDR foram planejadas para aderir às partições do IBGE, é possível obter a população oficial residente em cada partição correspondente.

4.3.1.1.c Cálculo do fator k para o local l : trata-se do fator de expansão constante para cada local. Assim sendo, as viagens realizadas pelos residentes do local l serão multiplicadas pelo fator k_l , de forma que:

$$k_l = \frac{\text{POP}_l}{\text{USU}_l}, l = 1..55 \quad (4.1)$$

Onde: POP_l é a população residente declarada pelo IBGE no local l dividida pela quantidade de usuários detectados como residentes do mesmo local l .

O fator de expansão fixo é obtido para todos os locais através da instrução SQL 4.3.1.1.c abaixo.

SQL 4.3.1.1.c _____ Versão completa em A.2.4.

```
_____ Fator de expansão populacional fixo _____
1  select PARTICAO_GEOGRAFIA, POP/count(ID) as k_l ...
2  from
3  residencia_presumida
4  inner join base_IBGE ....
```

A Figura 4.13 mostra a dispersão dos valores de k para as 55 unidades geográficas l que foram particionadas. Ou seja, trata-se da dispersão do *market share* local da operadora de telefonia. O *market share* global, nesse caso, é ≈ 5 . No capítulo de resultados, será mostrada a validação do fator de expansão fixo.

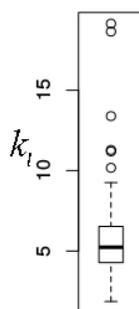


Figura 4.13: Dispersão de k_l fixo para as 55 unidades geográficas.

4.3.1.2 Fator de expansão adaptativo $k(l, d)$

O fator adaptativo busca atribuir maior peso aos usuários que, além de terem o local de residência presumido l , também registraram viagens no dia d , no trecho OD .

4.3.1.2.a Contar apenas usuários com residência presumida, idem à etapa 4.3.1.1.a.

4.3.1.2.b Contar as viagens detectadas $V(OD, l, d)$: trata-se basicamente da etapa para detecção de viagens através de duas chamadas consecutivas, $V(OD)$, a ser explicada em 4.4.1 ¹, contudo as viagens são segregadas entre os residentes de l no dia d .

4.3.1.2.c Totalizar as viagens de l no dia d : como base da probabilidade dos residentes de l percorrerem o trecho OD no dia d , é necessário agregar as viagens pelo local de residência presumida dos usuários, ou seja, $V(l, d)$, viagens dos residentes de l no dia d .

4.3.1.2.d Obter a população de, $POP(l)$, segundo o censo demográfico, idem ao item 1.b.

¹Por medida de simplificação, optou-se por focar a explicação apenas no fator de expansão, não detalhando a detecção de viagens nesse ponto. Esse detalhamento será posto na seção a seguir.

4.3.1.2.e Definir a probabilidade $P(OD|l)$: a probabilidade que ocorram viagens no trecho OD dado que essas viagens sejam percorridas pelos indivíduos residentes em l , pode ser definida como:

$$P(OD|l) = \frac{P(OD \cap l)}{P(l)} \quad (4.2)$$

A probabilidade $P(OD \cap l)$, pode ser obtida através da quantidade de viagens detectadas, de forma que:

$$P(OD \cap l) = \frac{V(OD, l)}{V} \quad (4.3)$$

Assim como, a probabilidade $P(l)$, pode ser obtida da seguinte forma:

$$P(l) = \frac{V(l)}{V} \quad (4.4)$$

Por medida de simplificação, o dia d foi retirado das expressões acima, que serviram apenas para embasar o cálculo da expectativa de viagens $\hat{V}(OD, d)$, que pode ser definido como:

$$\hat{V}(OD, d) = \sum_{l=1..55, d=1..362} \frac{V(OD, l, d)}{V(l, d)} \cdot POP(l) \quad (4.5)$$

A probabilidade que ocorram viagens no trecho OD dos residentes de l no dia d , é dividida pela probabilidade dos residentes de l realizarem viagens no dia d . Essa probabilidade é aplicada a todos os residentes $POP(l)$, expandindo a quantidade de viagens à população “real” (censo).

4.3.1.2.f Definir o fator adaptativo $k(OD, d)$: no contexto probabilístico definido acima, o fator de expansão adaptativo $k(OD, d)$, pode ser definido através da divisão entre a expectativa calculada de viagens e as viagens detectadas pelos dados de CDR:

$$k(OD, d) = \frac{\hat{V}(OD, d)}{V(OD, d)} \quad (4.6)$$

Para facilitar a visualização da dispersão dos valores de $k(OD, d)$, foi calculada a média do fator para todos os dias úteis, du , do ano, por OD , de forma que:

$$\overline{k(OD)} = \frac{\sum_{du=1..295} k(OD, du)}{295} \quad (4.7)$$

Assim sendo, a Figura 4.14 mostra a dispersão da média $\overline{k(OD)}$ para todos os dias úteis do ano, de onde o mínimo, a média e o máximo são 12, 36 e 178, respectivamente.

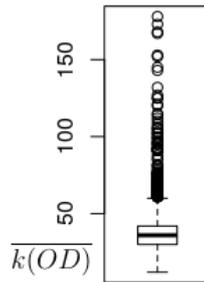


Figura 4.14: Dispersão de $\overline{k(OD)}$, média entre todos os dias úteis do ano, adaptado para os $55 \times 55 = 3025$ pares OD .

As metodologias a seguir farão uso dos fatores de expansão apresentados, os fatores serão referenciados apenas como “fator fixo k_l ” ou “fator adaptativo $k(OD, d)$ ”.

4.4 Matrizes de origem-destino quotidianas

Conforme visto na revisão bibliográfica (*Tabela 2.4*), a área de transportes adota diversos padrões de matrizes O-D para melhor compreender a demanda de mobilidade da população. A partir dos dados de CDR, do particionamento espacial, da inferência do local de residência dos usuários de telefonia, e considerando todos os modos e todos os motivos das viagens sem distinção, a metodologia apre-

sentada a seguir se propõe a calcular três tipos de matrizes O-D comuns à área de transporte: a) $N \times N$, de todos os locais para todos os locais; b) HBO, *home-based others*, ou seja, de *origem* casa para *destino* qualquer lugar, que no contexto desta pesquisa significa todos os lugares; e c) NHB, *non-home based*, ou seja, de *origem* qualquer lugar, **exceto casa**, para *destino* qualquer lugar.

Normalmente a Matriz $N \times N$ é a soma de todas as demais matrizes, por exemplo, $M_{N \times N} = M_{HBS} + M_{HBW} + M_{HBO} + M_{NHB}$. No entanto, sob a limitação do contexto desta pesquisa, a matriz $N \times N$ é apenas a soma entre HBO e NHB. Ou seja, $M_{N \times N} = M_{HBO} + M_{NHB}$. Além do mais, as matrizes devem ser compreendidas como *transientes*, assim como explicado na Fig. 2.9, na revisão bibliográfica.

A estimativa de matrizes O-D quotidianas (Fig. 4.15) é posterior ao domicílio presumido e anterior a apresentação dos resultados de mobilidade quotidiana.

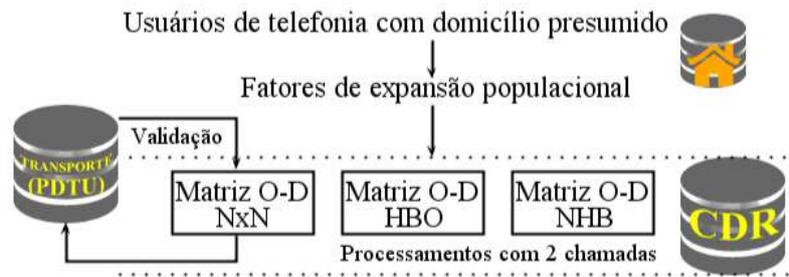


Figura 4.15: Fluxo metodológico das matrizes O-D quotidianas.

Para facilitar a compreensão, a primeira matriz a ser calculada será a $N \times N$, pois as demais matrizes HBO e NHB possuem uma etapa adicional. Exceto esse único passo adicional, o processo de cálculo é o mesmo para os três tipos de matrizes quotidianas.

4.4.1 Matriz O-D $N \times N$ (*de todos locais para todos locais*)

A matriz O-D $N \times N$ estima viagens sem distinção sobre o usuário ter partido de *origem* casa ou não. A seguir (Figura 4.16), a sequência de etapas adotadas para a obtenção da matriz $N \times N$. Observe que as etapas para construção da matriz $N \times N$ são idênticas às das matrizes HBO e NHB, inclusive o fator de expansão. A única

exceção está no passo c), de onde os contadores de viagens são distintos para os casos HBO e NHB, e serão vistos logo a frente na seção 4.4.2.

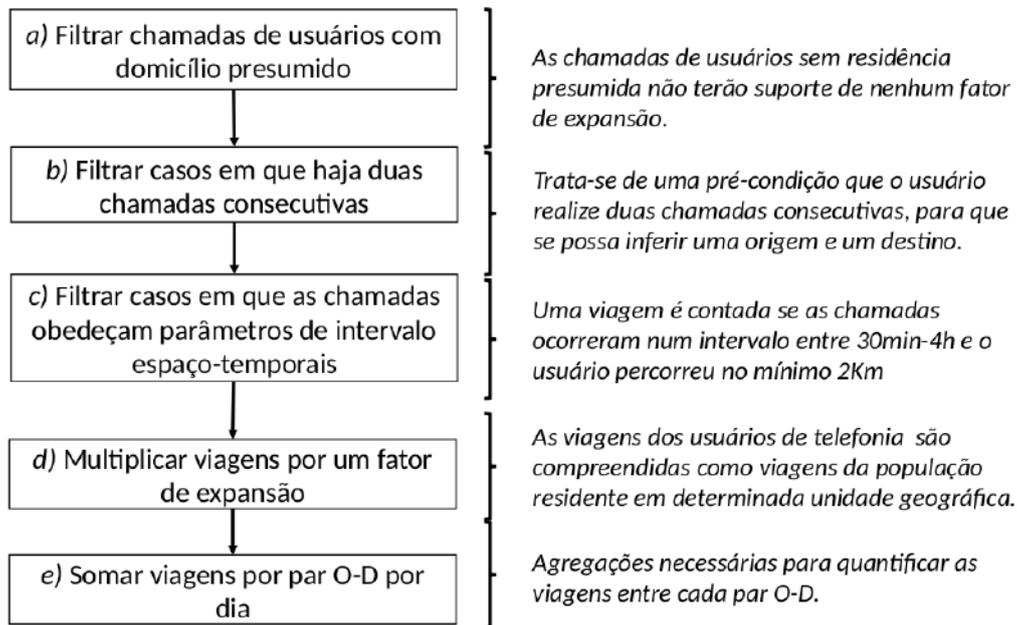


Figura 4.16: Fluxo de etapas para o cálculo geral de matrizes O-D diárias por dia, desde as chamadas realizadas por todos os usuários que tiveram a residência presumida até o resultado, que corresponde à matriz O-D $N \times N$.

Observe que as etapas a)-d) (Fig. 4.16) são responsáveis pela inferência dos deslocamentos em si. Na etapa d), a expansão das viagens em que o ponto de partida é diferente do local de residência (*presumivelmente trabalho ou escola*) é sempre feita em relação à **zona de domicílio do viajante**. A etapa e) trata da soma de viagens detectada entre todos os pares O-D por dia. O resultado será composto por n matrizes O-D, tanto quanto existam n dias.

As etapas a)-d) serão explicadas como itens separados, contudo a instrução SQL A.2.6, que está completa na seção de apêndice, executa tudo de uma só vez. Por vezes, alguma função da instrução SQL A.2.6 será citada sob numeração própria, para elucidar algum ponto específico, conforme a seguir.

a) Apenas usuários com domicílio presumido: trata-se da mesma forma já descrita em 4.3.1.1.a.

b) A detecção da viagem: o cálculo da matriz O-D quotidiana é dependente da realização de duas chamadas consecutivas de um mesmo ID. O fluxo “transiente” é inferido tradicionalmente dessa forma, conforme apresentado na revisão bibliográfica (Fig. 2.9).

A instrução responsável por satisfazer esta condição envolve o uso de uma função avançada, que compara duas linhas ordenadas em sequência. No caso, sob a ordenação de ID, DIA, HORA, a função `Conditional_change_event`, observa se houve mudança na posição geográfica (*latitude e/ou longitude*) da antena que atendeu a “próxima” chamada, conforme está destacado na instrução *SQL 4.4.1.b*, a seguir. Caso o ID não possua duas chamadas consecutivas, ele será ignorado.

SQL 4.4.1.b _____ Versão completa em A.2.6.

```

_____ Detecção de deslocamentos _____
1 ...
2         Conditional_change_event(b.latitude
3             || ','
4             || b.longitude)
5         OVER (
6             partition BY ddd_orig||id1
7             ORDER BY ddd_orig||id1,
8             dt_ini_atividade_rede||'_'||
9             hr_ini_atividade_rede)
10        AS DESLOCAMENTO ...

```

c) Intervalos espaço-temporais pré-definidos: uma viagem de um usuário de telefonia é determinada por uma distância percorrida sob um determinado tempo. Questões particulares da região de estudo precisam ser consideradas ao se estabelecer esse intervalo. Para efeito dos dados de CDR da RMRJ nesta pesquisa, um deslocamento é considerado uma viagem se foram “percorridos” no

mínimo 2km em no mínimo 30min; para efeito de limite máximo, foi considerado apenas o tempo de 4h.

Vale ressaltar que, a distância “percorrida” entre as chamadas consecutivas é a distância linear entre as torres de telefonia que atenderam a chamada. Ou seja, é uma distância aproximada, levando-se em conta que o usuário realiza chamadas sob a área de cobertura da antena. Essa questão de distância linear é relevante, constituindo uma limitação da metodologia, dada a topografia irregular do Rio de Janeiro, com diversas montanhas.

A parte do cálculo responsável por esta etapa está destacada no *SQL 4.4.1.c*, a seguir. Observe que foram usadas funções avançadas de geo-localização para obter a distância linear entre dois pontos; `ocoord` está para as coordenadas latitude/longitude da torre de origem, e `dcoord` está para a torre de destino. A função `Distance` retorna a distância em Km entre a origem e o destino. A função `Datediff()` retorna a quantidade de minutos entre uma chamada e outra.

SQL 4.4.1.c _____ Versão completa em A.2.6.

```
_____ Intervalo espaço-temporal _____  
1      ... Distance(occoord, dcoord) AS km_distance,  
2      Datediff(mi, hora, proxima_hora) AS minsbetween ...  
3      WHERE ocoord != dcoord //ORIGEM DIFERENTE DE DESTINO  
4      AND km_distance > 2 // Percorreu no min. 2 Km  
5      AND minsbetween > 30 // Intervalo temporal entre  
6      AND minsbetween < 240 // chamadas de 30min-4h ...  
_____
```

d) Aplicação do fator de expansão: sabendo-se que apenas as viagens de usuários com residência presumida foram consideradas, pode-se aplicar dois tipos de fator de expansão para se estimar as viagens da população: o fator fixo e o fator adaptativo. A seguir serão mostradas as instruções correspondentes a cada um dos casos.

Sendo que os estimadores já foram previamente calculados (ver seção 4.3.1), as instruções a seguir mostram onde deve-se multiplicar a quantidade de viagens de usuários, a fim de obter-se as viagens da população.

d.1) Instrução para estimar viagens através do k-fixo: a quantidade de viagens é multiplicada pelo k fixo através de `Count(*) * k_fixo`.

SQL 4.4.1.d.1 _____ Versão completa em A.2.6.

```
_____ Viagens por dia com k-fixo _____  
1 SELECT  
2 origem, destino, Count(*) * k_fixo as viagens_k_fixo, dia  
3 FROM ...  
4 ...INNER JOIN fator_k_l c ON c.PARTICAO_GEOGRAFIA=r.PARTICAO_GEOGRAFIA...
```

d.2) Instruções para estimar viagens através do k-adaptativo: note que o caso adaptativo precisa estar ligado aos residentes de l no dia d no trecho OD em que ocorreram as viagens. Primeiramente, precisa-se obter a quantidade total de viagens realizadas por l no dia d , trata-se do equivalente à Equação 4.4 da Seção 4.3.1, depois é necessário calcular a quantidade de viagens segundo a Equação 4.5, o que equivale à instrução: `ceil((a.viagens/b.viagens)*(c.pop_censo)) as viagens_adaptativas` Esses dois cálculos constam na instrução A.2.5 do Apêndice.

e) Soma de viagens por O-D por dia: trata-se da finalização do cálculo, mostrando a estimativa de viagens entre todos os pares O-D da RMRJ, para os n dias disponíveis na base de CDR.

As agregações por dia podem revelar diferentes comportamentos de mobilidade da população por tipos de dia, tais como dias úteis, fins-de-semana ou feriados prolongados.

4.4.2 Matrizes O-D HBO e NHB

Conforme explicado no início desta seção, as matrizes HBO e NHB possuem uma etapa adicional em relação à matriz NxN. A única função desta etapa é averiguar se o local de origem da viagem era ou não, o mesmo do que o local de residência presumida do usuário. Essa função está destacada, a seguir, na Figura 4.17, pelas sub-etapas **c.1** e **c.2**.

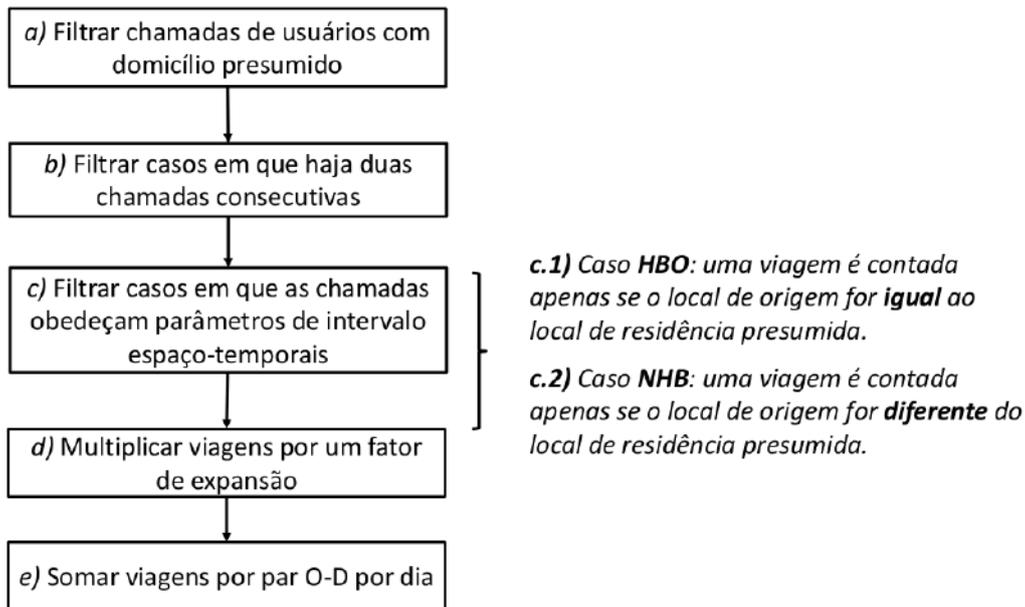


Figura 4.17: Fluxo de etapas para o cálculo geral de matrizes O-D quotidianas por dia, destaque para as condições que calculam as matrizes HBO e NHB, c.1 e c.2 respectivamente.

As etapas principais apresentadas na Figura 4.17, são basicamente as mesmas da Figura 4.16; por esse motivo, serão descritas a seguir, apenas as sub-etapas **c.1** e **c.2**.

Para facilitar a compreensão, pode-se comparar os passos a seguir com a etapa **c** da seção 4.4.1, que resulta na matriz NxN.

c.1 Matriz HBO: apenas viagens cujo ponto de partida é **igual** ao local de residência do usuário que realizou a viagem, essa nova condição é estabelecida através da cláusula `...AND (id, origem) in (select ID, PARTICAO_GEOGRAFIA....`

A instrução *SQL 4.4.2.c.1* que monta a matriz HBO é embasada na *SQL 4.4.1.c*, a partir da matriz NxN.

SQL 4.4.2.c.1 _____ Versão completa em A.2.6.

Matriz HBO

```
1  ...
2  WHERE ocoord != dcoord AND km_distance > 2 AND minsbetween > 30
3      AND minsbetween < 240/*abaixo->condição HBO*/
4  AND (id,origem) in
5  (select ID, PARTICAO_GEOGRAFIA from residencia_presumida) ...
```

c.2 Matriz NHB: apenas viagens cujo ponto de partida é **diferente** do local de residência do usuário que fez a viagem, essa nova condição é estabelecida através da cláusula `...AND (id,origem) NOT in (select ID, PARTICAO_GEOGRAFIA....`

Da mesma forma que a matriz HBO, a instrução *SQL 4.4.2.c.2* que monta a matriz NHB é embasada na *SQL 4.4.1.c*, a partir da matriz NxN.

SQL 4.4.2.c.2 _____ Versão completa em A.2.6.

Matriz NHB

```
1  ...
2  WHERE ocoord != dcoord AND km_distance > 2 AND minsbetween > 30
3      AND minsbetween < 240/*abaixo->condição NHB*/
4  AND (id,origem) NOT in
5  (select ID, PARTICAO_GEOGRAFIA from residencia_presumida) ...
```

As matrizes O-D HBO e NHB encerram as metodologias de matrizes O-D quotidianas (ou ordinárias) abordadas nesta pesquisa. A seguir, será apresentada uma abordagem de matriz O-D extraordinária, que não é tão comum à área de transportes.

4.5 Matriz de origem-destino extraordinária

A partir das medições espaço-temporais possíveis com os dados de CDR, viu-se a oportunidade de abordar um tipo de matriz O-D extraordinária, dada a eventos pontuais, com delimitações espaço-temporais próprias. O posicionamento metodológico do cálculo de matriz O-D extraordinária pode ser visto na Figura 4.18, a seguir.



Figura 4.18: Posicionamento metodológico da matriz O-D extraordinária.

A abordagem quotidiana pode ser compreendida sob um comportamento regular ou ordinário dos usuários de telefonia durante um ano, contudo padrões esporádicos ou extraordinários de viagens estão diluídos e ocultos dentro das matrizes quotidianas. Esses padrões geralmente são compreendidos como **eventos**, e podem ser explorados com maior profundidade. Será apresentada a seguir uma metodologia para esse fim.

Conforme visto na revisão bibliográfica (*Seção 2.3*), o conceito de evento é amplo, abrangente e de certa forma popular, já que muitas pessoas são atraídas por eventos e associam-nos a momentos de lazer, de cultura e até mesmo de conhecimentos gerais.

O ponto de partida da matriz O-D extraordinária é o mesmo das matrizes quotidianas, ou seja, a base de usuários e seus respectivos locais de domicílio presumido. A partir de então, os procedimentos mudam consideravelmente, ver Fig. 4.19; uma delimitação espaço-temporal é definida livremente, o fator de ex-

pansão é adaptativo e delimitado ao intervalo do evento, e os processamentos levam em conta apenas **uma chamada telefônica**, ao invés de duas consecutivas (*esquema quotidiano*). A seguir, uma descrição das etapas elencada na Figura 4.19.

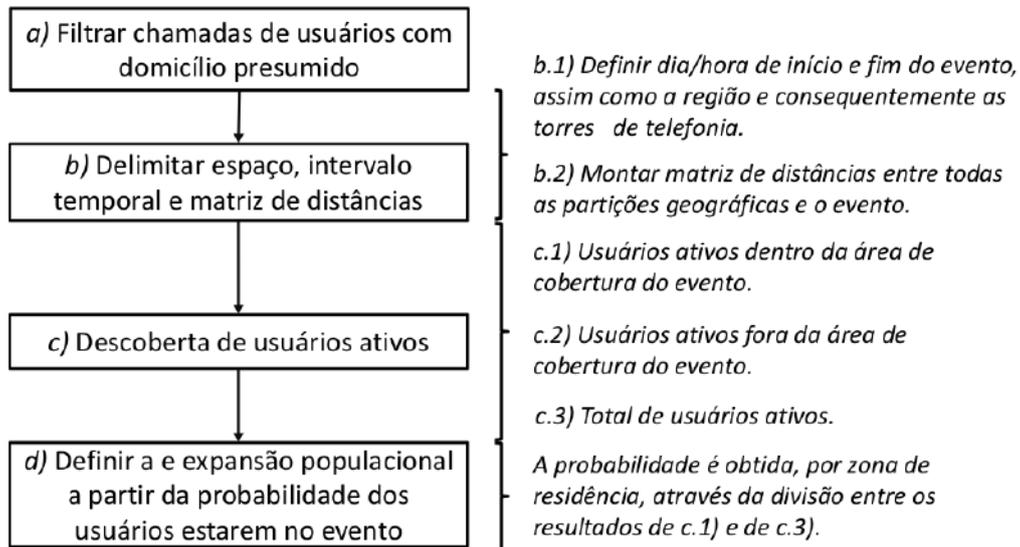


Figura 4.19: Fluxo de etapas para o cálculo da matriz O-D extraordinária.

a) Apenas usuários com domicílio presumido: trata-se da mesma forma já descrita no item **1.a** em 4.3.1.

b.1) Delimitação espaço-temporal: a delimitação temporal trata-se pura e simplesmente de uma data e hora de início e uma data e hora de fim do evento. Já o processo de delimitação espacial requer maiores detalhes, conforme será mostrado seguir.

A compreensão da delimitação espacial é facilitada ao recordar-se da definição de particionamento espacial apresentada na Seção 4.2 (Fig. 4.7), onde cada unidade geográfica foi mostrada como uma composição de mais de uma torre de telefonia móvel. O caso com a delimitação para eventos se dá de forma similar, contudo ao invés de aproximar uma região a um subdistrito, a região do evento pode ser aproximada livremente de acordo com a área evento.

A delimitação espacial é a *etapa-base* para o cálculo da matriz O-D extraordinária. Diferente das demais etapas, resolvidas por códigos SQL, a delimitação

espacial de um evento é um processo “manual”, já que qualquer conjunto de torres de telefonia t pode ser escolhido livremente dentre todas as torres T da área de estudo, independente das partições geográficas.

Deve-se compreender a área de cobertura do evento a partir dos dados divulgados pela própria organização. Para facilitar o entendimento dessa abordagem, tomemos o exemplo do Reveillon em Copacabana, sabe-se que o evento ocorreu na extensão da orla, assim sendo, recomenda-se que o conjunto de torres de telefonia t seja composto por todas as torres que cobrem a orla, **mais** uma margem para além da orla. Essa margem é necessária por dois motivos: a) muitas pessoas podem ter realizado chamadas em regiões próximas ao centro do evento, ou b) as antenas sobrecarregadas podem ter a carga “balanceada” com antenas próximas. Detalhes técnicos sobre os mecanismos das torres de telefonia podem ser vistos em [17] e [71].

Existe a possibilidade de que essa informação sobre a cobertura de um evento seja declarada ou “percebida” em suas reais dimensões após sua ocorrência, daí a necessidade de um grau de liberdade maior do que o esquema cotidiano.

Uma vez descoberta a região ρ do evento (com a devida margem), a escolha das torres t pode ser feita segundo a lógica do Algoritmo 1. Essa operacionalização pode ser facilitada por alguma ferramenta de manipulação de mapas (tipo QGIS), onde seja possível visualizar a sobreposição de camadas, para poder escolher visualmente as torres, a partir da região ρ .

Conforme visto na Seção 4.3.1, a matriz O-D cotidiana considera os fatores de expansão populacional na forma fixa e adaptativa. Para a matriz O-D extraordinária viu-se a necessidade de criar um fator de expansão adaptativo e delimitado ao intervalo do evento, conforme será visto a seguir.

b.2) Calcular distâncias entre todas as partições geográficas e o evento: a distância a percorrer é um dos fatores de impedância conhecidos pela área de transportes.

Algoritmo 1: Escolher as torres de telefonia t que atenderam um determinado evento na região ρ .

Entrada: Todas as torres de telefonia T da área de estudo.

Parâmetros: Região geográfica ρ em que ocorreu o evento.

Saída: Lista de torres de telefonia $t \in T$ que cobriram o evento.

```
1  $t = \text{null};$ 
2  $T = \text{todas as torres disponíveis};$ 
3  $k = 0; i = 0;$ 
4 para  $t_i \in T$  faça
5   | se  $t_i \subset \rho$  então
6   |   |  $t_k.LatLong = t_i.LatLong;$ 
7   |   |  $k = k + 1;$ 
8   | fim
9   |  $i = i + 1;$ 
10 fim
```

Ao medir a distância percorrida através das vias de acesso, espera-se calcular a probabilidade de uma pessoa residente no local l de dirigir-se ao evento ϵ . Observe que, para a finalidade de “medir a impedância”, a distância linear não seria viável, por esse motivo que foi empregada uma API² própria do Google Maps, chamada *Distance Matrix*³.

A API recebe por parâmetro duas coordenadas geográficas, uma de origem e outra de destino, e retorna a distância em metros a percorrer pelas vias de menor tempo estimado. Observe, por exemplo na Figura 4.20, a distância calculada entre um ponto no bairro Cosme Velho e outro em Botafogo, um 960m distante do outro, contudo 5,4Km a percorrer se for considerada a distância entre as vias de acesso, pois há uma montanha entre os dois pontos.

Esta etapa é responsável por salvar as distâncias das vias de acesso entre as coordenadas de centro de todas as partições geográficas de estudo e o centróide da região do evento. Esses valores serão usados em passos posteriores para estimar a probabilidade de um residente de determinada região ir a um evento específico.

²Application Program Interface.

³Site da API <https://developers.google.com/maps/documentation/distance-matrix/>, visto em 24 de fevereiro de 2018.

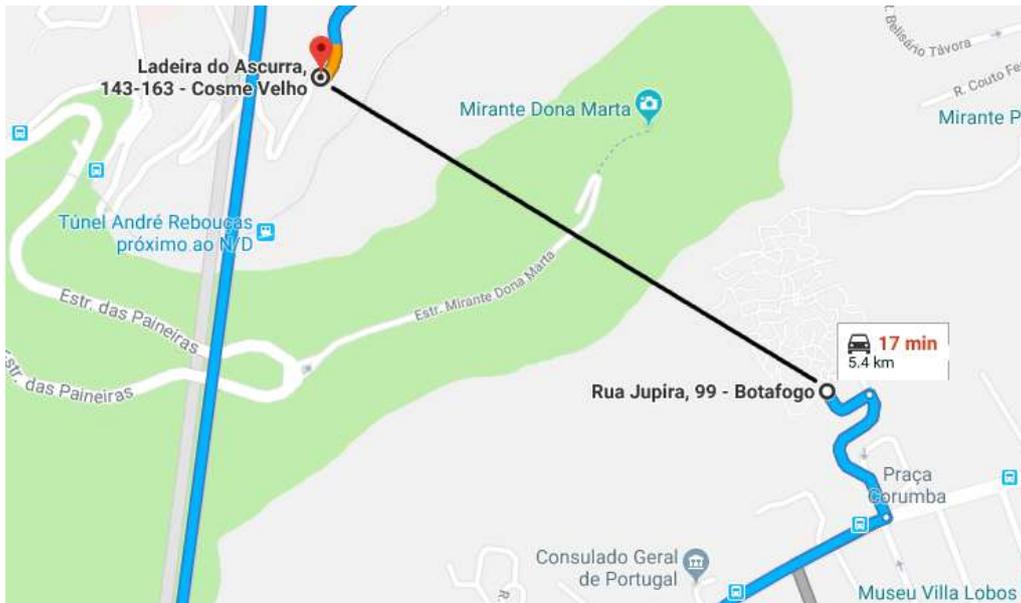


Figura 4.20: Diferença entre a distância linear (linha preta) e a distância pelas vias de acesso (azul). Fonte: Google Maps.

c) **Descoberta de usuários ativos:** além da pré-condição do usuário possuir o domicílio presumido, é necessário que o usuário de telefonia seja detectado como “ativo”, dentro ou fora da região de cobertura do evento, conforme a seguir.

Em continuidade ao passo anterior, as torres de telefonia t que foram escolhidas manualmente, precisam estar organizadas em formato tabular (*linhas e colunas*), para a evolução das etapas posteriores. Um exemplo dessa organização pode ser vista na Figura 4.21, onde é visível que um evento deve possuir $1 \dots N$ posições de latitude/longitude (torres t), e cada torre t pode atender $0 \dots N$ chamadas (CDR).

A partir do modelo ER apresentado na Figura 4.21, é possível atender às sub-etapas abaixo.

c.1) *Descoberta de usuários dentro da região de cobertura do evento:* o evento possui uma data/hora de início e fim, abrangendo N torres t (em EVENTO_LATLNG), e cada torre pode atender N chamadas. Um usuário de telefonia é considerado “dentro” do evento quando realiza uma chamada que é atendida por uma torre que abrange o evento.

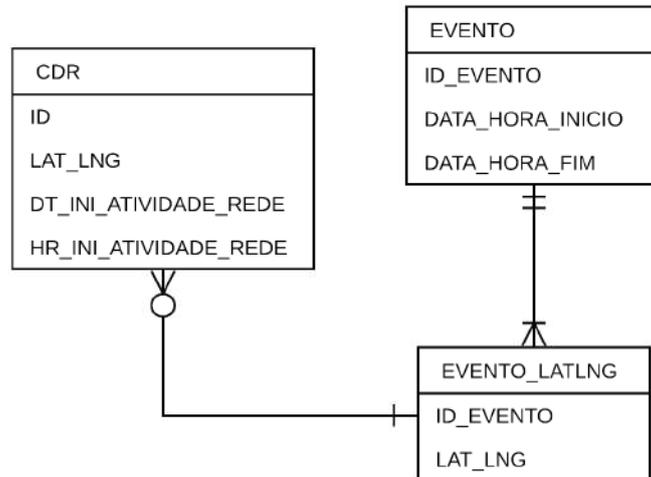


Figura 4.21: Modelagem Entidade/Relacionamento (ER) para eventos.

A instrução *SQL 4.5.c.1*, a seguir, traz todos os IDs que registraram atividade no evento de Reveillon de 2014, por exemplo.

SQL 4.5.c.1 _____ Versão completa em A.2.6.

Usuários no evento

```

1 create table EVENTO_REVEILLON2014_DENTRO as
2 select ID , LATITUDE_LONGITUDE .... where ....
3 and b.ID_EVENTO = 'REVEILLON2014'
4 group by ID, a.LAT_LNG;
  
```

c.2) Descoberta de usuários *fora* da região de cobertura do evento: a instrução é similar a anterior, contudo deve-se buscar a porção complementar do conjunto formado por `EVENTO_REVEILLON2014_DENTRO`. Esse complemento é obtido através da instrução `FULL OUTER JOIN` entre `CDR` e `EVENTO_LATLNG` e as instruções `a.LAT_LNG is null and b.LAT_LNG is null`, no *SQL 4.5.c.2*.

SQL 4.5.c.2 _____ Versão completa em A.2.6.

_____ Usuários fora do evento _____

```
1 create table EVENTO_REVEILLON2014_FORA as
2 select a.ID , a.latlng
3 from CDR a FULL OUTER JOIN EVENTO_LATLNG b on
4 where .... and b.ID_EVENTO = 'REVEILLON2014'
5 and a.LAT_LNG is null and b.LAT_LNG is null
6 group by ID, a.LAT_LNG;
```

c.3) *Total de usuários ativos:* os conjuntos de dentro e de fora da região ρ do evento, podem ser unidos através da expressão UNION ALL, conforme SQL 4.5.c.3 abaixo.

SQL 4.5.c.3 _____

_____ Total de usuários ativos durante o evento _____

```
1 create table EVENTO_REVEILLON2014_TOTAL as
2 select ID , latlng from EVENTO_REVEILLON2014_DENTRO
3 UNION ALL
4 select ID , latlng from EVENTO_REVEILLON2014_FORA;
```

d) **Obter o fator de expansão adaptativo, delimitado ao evento:** o fator de expansão adaptado ao evento é similar ao apresentado para a matriz O-D quotidiana, contudo é necessário compreender que o evento é visto como destino único de todas as origens, e que a presença dos “visitantes” sob a delimitação espaço-temporal do evento é obtida através de **uma** chamada telefônica.

A probabilidade de que os residentes de l sejam visitantes do evento ϵ , $P(l|\epsilon)$, pode ser compreendida como:

$$P(l|\epsilon) = \frac{P(U(l \cap \epsilon))}{P(U(\epsilon))} \quad (4.8)$$

Onde, $P(U(l \cap \epsilon))$ é a probabilidade de usuários residentes em l estarem no evento ϵ , sob o intervalo temporal do próprio evento. $P(U(\epsilon))$, refere-se a probabilidade de todos usuários que realizaram no mínimo uma chamada sob o intervalo temporal do evento, independentemente de onde estavam.

As probabilidades podem ser extraídas dos dados de CDR da seguinte forma:

$$P(U(l \cap \epsilon)) = U(l, \epsilon) \quad (4.9)$$

$$P(U(\epsilon)) = U(\epsilon) \quad (4.10)$$

Onde, $U(l, \epsilon)$ é a quantidade de usuários residentes em l que realizaram no mínimo uma chamada na região delimitada do evento ϵ , sob o intervalo temporal do próprio evento. $U(\epsilon)$, refere-se a todos os usuários que realizaram no mínimo uma chamada sob o intervalo temporal do evento, independentemente de onde estavam.

A expectativa de visitantes (*expandida*) de l ao evento ϵ , $\hat{\Omega}(l, \epsilon)$, pode ser dada por:

$$\hat{\Omega}(l, \epsilon) = \frac{U(l, \epsilon)}{U(\epsilon)} \cdot POP(l) \quad (4.11)$$

Onde, $POP(l)$, refere-se a população oficial (censo) de l . Observe que optou-se por usar a letra Ω ao invés de V , para não confundir com o conceito de Viagens, explicado para a matriz O-D quotidiana. Já que, no cálculo de pessoas em eventos são consideradas “**visitas**” ao local do evento, e não *viagens*.

Assim sendo, o fator de expansão, aplicado ao residentes de l e delimitado ao evento ϵ , pode ser definido como:

$$k(l, \epsilon) = \frac{\hat{\Omega}(l, \epsilon)}{U(l, \epsilon)} \quad (4.12)$$

Concluída a definição da expansão, observe no *SQL 4.5.e* que a distância a percorrer pelas vias de acesso faz parte do resultado ("DISTÂNCIA (m) "). A relação entre probabilidade e distância será explorada no capítulo de resultados.

SQL 4.5.e _____ Versão completa em A.2.6.

Quantidade de visitantes em eventos

```
1  select
2  nm_grao as "LOCAL",
3  nm_municip as "MUNICÍPIO",
4  zona as "ZONA",
5  pop_censo as "POPULAÇÃO",
6  pop_usu as "USUÁRIOS",
7  Distance_meters as "DISTÂNCIA(m)",
8  (USR_IN/USR_TOT) * POP_IBGE as "VISITANTES" ...
```

A principal diferença entre os resultados do esquema ordinário e extraordinário, é que o primeiro leva em conta a realização de *duas chamadas telefônicas* (uma viagem), enquanto o último considera apenas *uma chamada*, equivalente a detecção de uma visita, assumindo o local de residência presumida como pretenso origem.

Enfim, foi apresentada a área de estudo, a descrição da base de dados, e as instruções desenvolvidas para atingir os resultados da pesquisa. Também foram mostrados fatores de expansão populacionais, levando-se em conta os dados do censo demográfico. Sobre os eventos extraordinários, será proposto na seção de resultados um “direcionamento” para validação, que envolve a densidade de pessoas/m².

Capítulo 5

Resultados de matrizes O-D quotidianas

Conteúdo

5.1	Discussão sobre o particionamento espacial	97
5.2	Validação dos fatores de expansão populacional	99
5.3	Validação da quantidade de viagens	101
5.4	Identificação de padrões de mobilidade	103
5.5	Detecção de mudança em padrões de mobilidade	116

A seguir serão apresentados os resultados de matrizes O-D quotidianas obtidos através dos dados de CDR, seguindo as instruções SQL e as metodologias explicadas no capítulo anterior. No entanto, antes dos resultados é necessário estabelecer as condições de validade da metodologia, de forma que, ao longo da exposição serão discutidas e contextualizadas essas validações, a começar por uma discussão sobre o particionamento espacial.

5.1 Discussão sobre o particionamento espacial

Seguindo a ordem apresentada na metodologia, o primeiro questionamento é sobre a granularidade de estudo, ou do particionamento geográfico adotado: *a)* por que a unidade geográfica da área de cobertura de cada torre de telefonia não foi

adotada como unidade geográfica? Ou *b)* por que não foi adotada uma unidade geográfica maior, do tipo macrozonas de tráfego?

A validade do particionamento espacial está implicitamente condicionada à validade das etapas posteriores da metodologia, já que a unidade geográfica é a base para todas as demais etapas, influenciando diretamente os resultados, de forma especial a detecção de domicílio.

Seguindo a lógica de que o particionamento espacial influencia a detecção de domicílio, pode-se comparar os resultados deste estudo com os de outra pesquisa que usou os mesmos dados e metodologia similar de domicílio presumido, contudo adotou o particionamento espacial sob o nível de polígono de voronoi das torres de telefonia, ou seja, item *a)* do primeiro parágrafo.

A partir dos mesmos dados de CDR usados nesta pesquisa, em recente estudo, JOSÉ [18] fez inferências sobre a quantidade de pessoas presentes no Reveillon e no bloco popular de carnaval “Cordão do Bola Preta”. Os resultados mostraram um fator de expansão populacional fixo médio de $\bar{k} \approx 22$, ou seja, cada usuário de telefonia com residência presumida foi equivalente a 22 pessoas. Comparativamente, esta pesquisa obteve um valor médio $\bar{k} \approx 5$ (ver Seção 4.3.1). Considerando que, quanto maior o valor médio de \bar{k} , menor foi a taxa de detecção (e vice-versa), esta pesquisa obteve uma taxa de detecção de usuários 4x maior, aproximadamente.

Uma possível explicação para o fenômeno descrito acima, é que nas regiões com maior densidade de antenas, é possível que chamadas realizadas de um mesmo local sejam atendidas ora por uma antena, ora por outra, dadas as regiões de limite de cobertura entre as torres¹. Outra explicação plausível é que o usuário possa realizar muitas chamadas em locais próximos de casa, tais como locais de comércio, padarias, bares e supermercados.

Em contrapartida às unidades geográficas como unidade de torre de telefonia, alguém pode querer adotar um agrupamento que contemple regiões que podem

¹Para detalhes técnicos sobre as antenas, ver XAVIER [17] e DOYLE [71].

ser compostas de várias cidades, como é o caso das macrozonas de tráfego (item *b*) do primeiro parágrafo). A desvantagem, nesse caso, seria a imprecisão por conta da abrangência geográfica.

Enfim, a baixa taxa de detecção de domicílio presumido ao se usar uma unidade geográfica sob o nível de torres de telefonia, pode ser compensada com uma maior precisão geográfica daqueles que foram detectados. É preciso aplicar o viés que seja mais próximo dos objetivos da pesquisa.

A seguir será apresentada a validação para o fator de expansão populacional fixo, considerando as unidades geográficas conforme apresentado na Seção 4.2, ou seja, subdistritos para a cidade do Rio de Janeiro e as próprias cidades nas demais regiões.

5.2 Validação dos fatores de expansão populacional

A validação do cálculo de domicílio presumido foi feita com os dados do censo do IBGE de 2010, como mostra a Figura 5.1, de onde pode-se observar um coeficiente de determinação de 0,87, obtido entre os usuários de telefonia móvel com local de residência presumida e a população oficial (censo) respectiva ao mesmo local.

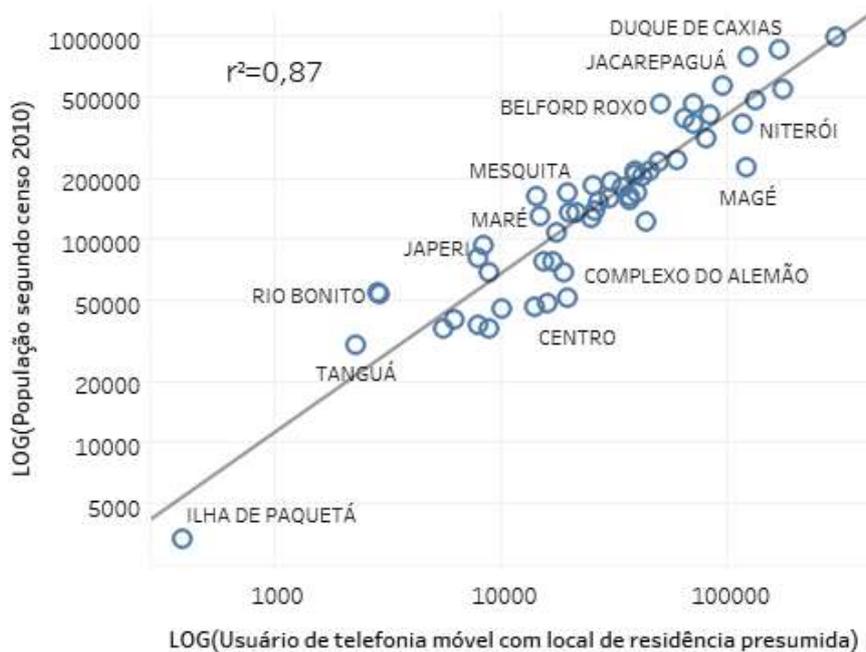


Figura 5.1: Usuários de telefonia detectados em 2014 nas unidades geográficas de estudo *vs.* população declarada pelo censo 2010 para as mesmas unidades.

Conforme explicado na metodologia, o fator de expansão k-fixa, é unicamente embasado na detecção de residência dos usuários de telefonia, motivo pelo qual é razoável a comparação com os dados do censo (conforme Figura 5.1). Já o k-adaptativo foi concebido para que a atividade de viagens dos usuários num determinado dia e sob determinado par O-D influenciasse na correção populacional. Por esse motivo, a validação do fator adaptativo precisa ser realizada diretamente com a quantidade de viagens.

A título de validação da quantidade de viagens, foi usada uma base oficial da área de transportes da RMRJ em comparação com a quantidade obtida através dos dados de CDR. As condições de validação para todos os resultados de matrizes O-D quotidianas serão apresentadas a seguir.

5.3 Validação da quantidade de viagens

Conforme visto na revisão bibliográfica, os relatórios da área de transporte estimam viagens para um único dia útil comum, ao passo que a partir dos dados de CDR, é possível de se estimar viagens para todos os dias disponíveis na base.

Para tornar-se viável uma comparação entre a quantidade de viagens estimada pelo relatório de transportes da RMRJ (PDTU 2013) e a dos CDR, foram necessários alguns processamentos até que ambas as bases de dados estivessem sob a mesma escala.

O relatório do PDTU de 2013 apresenta os resultados sob o nível de macrozonas, que correspondem a **conjuntos** de subdistritos e cidades, já o PDTU está disponível sob o nível de zonas de tráfego. Foi necessário agregar conjuntos de zonas em subdistritos e cidades, igualando as unidades geográficas do PDTU com as próprias unidades geográficas do estudo.

Além da equivalência de particionamento espacial, o PDTU possui diversas categorias de viagens, se foram feitas de forma individual, coletiva, se foi no pico da manhã, da tarde, etc. Foi necessário somar todas essas categorias, obtendo-se uma quantidade total de viagens entre cada par O-D, que nesse ponto já correspondia às unidades geográficas O-D próprias do estudo.

A partir dos dados de CDR é possível estimar a quantidade de viagens realizadas entre todos os pares O-D para todos os dias de estudo. Para que a quantidade de viagens entre o PDTU e CDR fosse comparada sob um mesmo “princípio”, foi calculada a média de viagens detectadas via CDR a partir de todos os dias úteis do ano, tanto para o k-fixa quanto para o k-adaptativo.

Houve casos em que o CDR detectou viagens entre pares O-D que não foram relatados no PDTU, e vice-versa. Para efeito de comparação, esses casos foram ignorados, fazendo valer apenas a comparação entre pares O-D em que as viagens de todas as fontes fossem maiores que 0.

Seguindo todas as condições de validação explicadas acima, obteve-se uma relação linear entre a quantidade de viagens inferida pelo PDTU e as inferidas

pelo CDR sob a média dos fatores k-fixo e k-adaptativo em dias úteis (Figura 5.2), com coeficiente de determinação de 0,81 e 0,78, respectivamente.

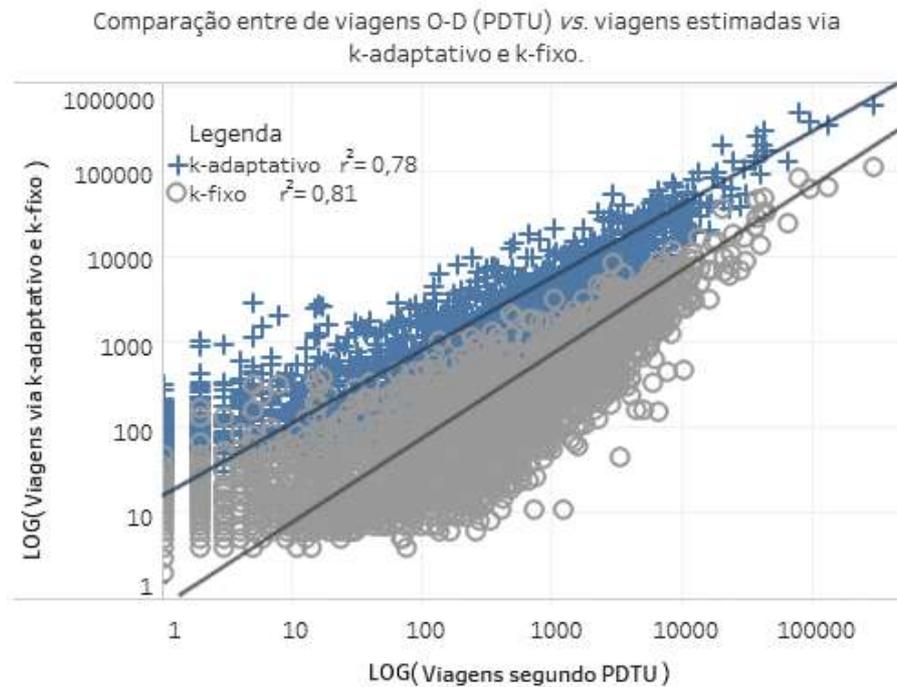


Figura 5.2: Comparação entre a quantidade de viagens inferida pelos fatores de expansão populacionais k-fixo e k-adaptativo vs. PDTU.

A diferença entre os fatores reside na escala dos valores. Se consideramos a média de viagens entre todos os pares O-D e todos os dias úteis, o k-adaptativo estima uma quantidade de viagens 6x maior do que o k-fixo. De forma geral, pode-se observar que o k-adaptativo superestimou a quantidade de viagens em relação ao PDTU.

A Figura 5.3 foi montada a título de verificação da coerência entre a metodologia de estimativa de viagens via k-fixo e via k-adaptativo. Pode-se observar os mesmos indicadores já usados nos gráficos anteriores, ou seja, a média de viagens entre os trechos *OD* de todos os dias úteis do ano.

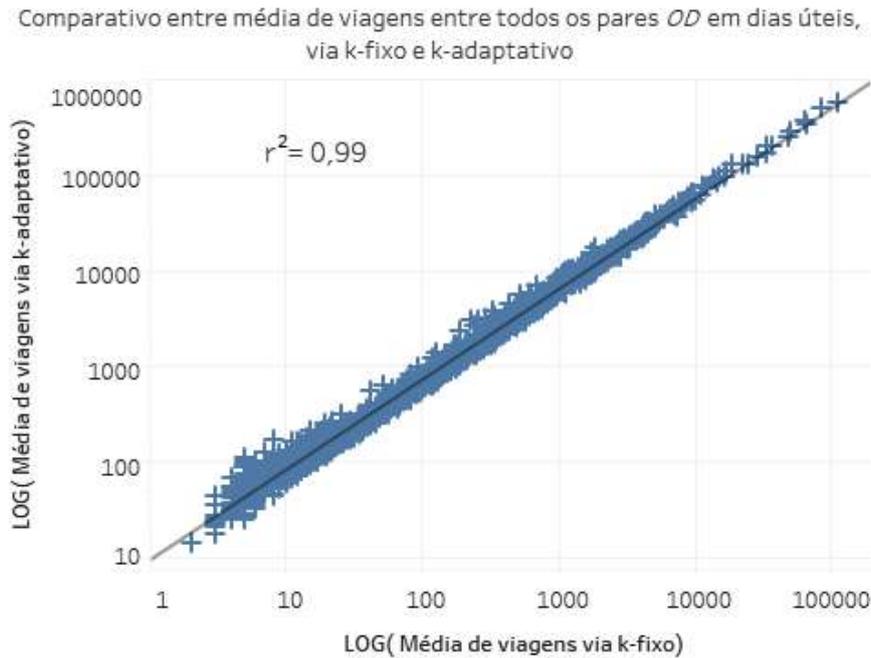


Figura 5.3: Comparação entre a média de viagens entre todos os pares *OD* em dias úteis, tanto via k-fixe quanto via k-adaptativo.

Enfim, os dois fatores de expansão populacionais apresentados mostraram coerência entre si, contudo o fator k-adaptativo superestima a quantidade de viagens, elevando-se em relação ao k-fixe e ao próprio PDTU. Em contrapartida, o fator k-fixe parece subestimar a quantidade de viagens. Por esse motivo, alguns resultados em relação à matrizes O-D quotidianas, a seguir, serão apresentados sob o efeito comparativo dos dois fatores, de onde espera-se evidenciar essa diferença de escala.

5.4 Identificação de padrões de mobilidade

Dada a abrangência dos dados de CDR, a análise dos resultados por dia revelou padrões de mobilidade bem característicos, alguns mais marcados e outros mais sutis. Sendo que a área de estudo contempla 3025 pares O-D (55 x 55), esses padrões foram descobertos empiricamente através de perguntas simples, tais como:

a) qual seria a região que atrai grande quantidade de pessoas ao trabalho? b)

qual seria o destino de lazer mais atrativo para residentes de um determinado local? c) dados os meios de transporte de massa disponíveis, quais seriam os destinos mais atrativos para lazer?

Em termos de padrões que se repetem ao longo do ano, foram encontrados dois tipos de padrões quotidianos: Casa-Trabalho e Casa-Lazer. O fator de expansão k-fixo foi definido como medida padrão para a apresentação dos resultados, contudo serão apresentados alguns *insights* revelados pela comparação entre valores obtidos através do k-fixo e do k-adaptativo. Os resultados desses padrões serão apresentados a seguir.

Casa - Trabalho

Tratando-se da RMRJ, o centro da cidade do Rio de Janeiro é um dos maiores polos atratores de empregos. Partindo-se desse princípio e observando-se os resultados, há um padrão regular e consistente em mostrar o centro como destino de trabalho para residentes de diversas áreas. Isso ocorre principalmente em dias úteis, evidenciando uma característica típica de trabalho no destino.

Através da análise das três matrizes quotidianas desenvolvidas nesta pesquisa, pode-se descobrir se as pessoas saíram de suas casas (HBO), ou se determinada região é usada apenas como corredor (NHB). Essas características podem ser vistas na Figura 5.4, onde os subdistritos de Copacabana, Botafogo e Lagoa (Zona Sul) foram considerados como origem, e o centro como destino. Observa-se que a maior parte das viagens Zona Sul x Centro não são realizadas por moradores da Zona Sul, pois as viagens NHB oscilam numa medida $\approx 2x$ maior do que as viagens HBO. Observe também que, as regiões do topo dessa série temporal (Fig. 5.4) indicam os dias úteis (Seg-Sex).

Vale lembrar que, segundo explicado na metodologia, Matriz $N \times N =$ Matriz NHB + Matriz HBO.

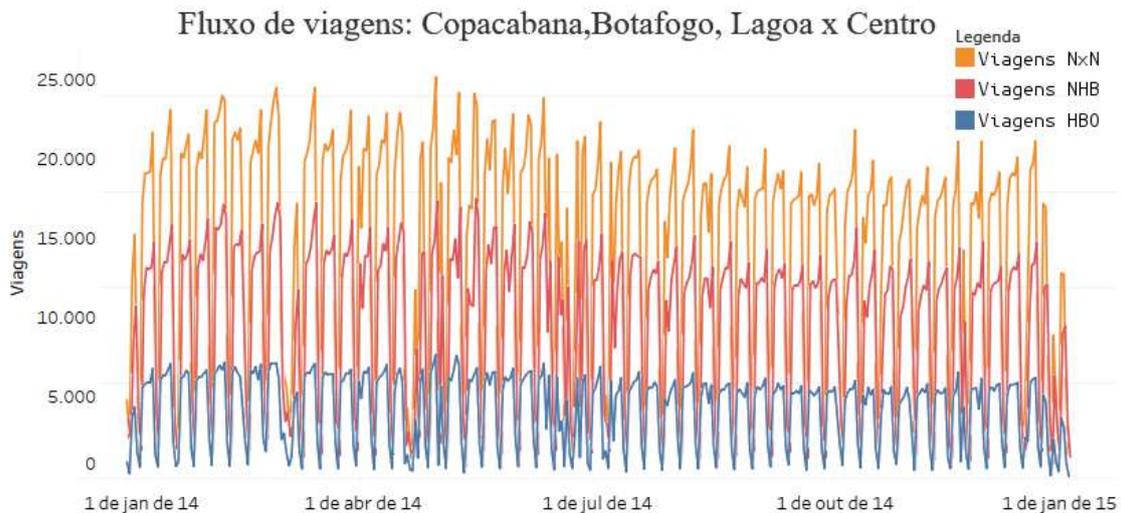


Figura 5.4: Quantidade de viagens zona sul x centro (k-fixa), visto através das matrizes NxN, NHB e HBO.

A exceção de alguns feriados prolongados de 2014, como Páscoa e Carnaval, Natal e Ano Novo, ou a própria Copa do Mundo em Julho, o padrão casa-trabalho se mostra regular, com uma maior proporção de viagens em dias úteis e menor em dias de descanso. A Figura 5.5 confirma o comportamento já destacado na Figura 5.4, que a mobilidade da Zona Sul ao Centro é maior em dias úteis e menor nos fins-de-semana, tendo um pico nas sextas-feiras.

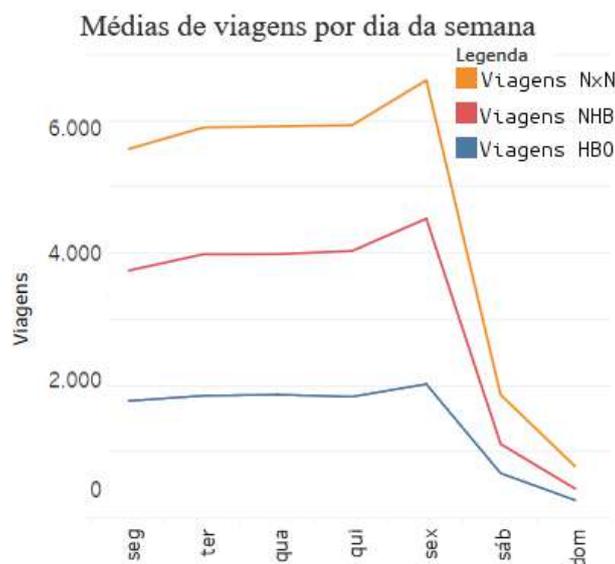


Figura 5.5: Padrão semanal Casa-Trabalho.

Uma das principais limitações da metodologia é a dependência do usuário realizar uma chamada para ser “detectado”. Sendo assim, é possível que esse pico nas sextas-feiras seja apenas uma tendência de as pessoas ligarem mais nesse dia, conseqüentemente ampliando o universo de usuários “detectados”. Essa tendência pode ser confirmada se for observada (Fig. 5.6) a média de chamadas realizadas por dia da semana, de todos os usuários com residência presumida.

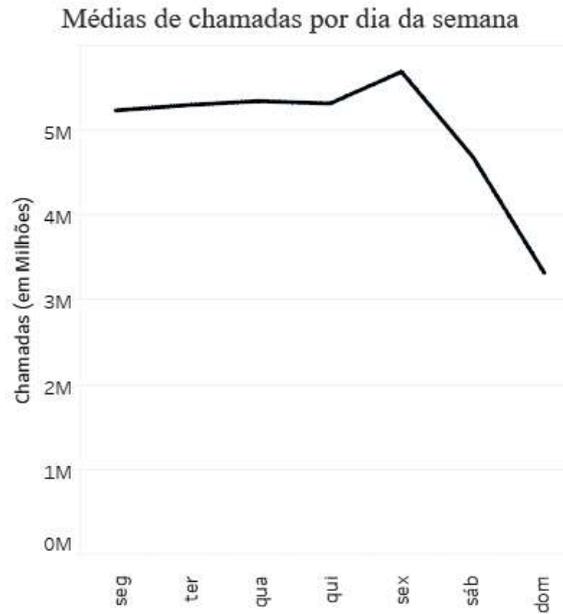


Figura 5.6: Média de chamadas por dia da semana de todos usuários com residência presumida.

A Figura 5.7 foi montada a título de comparação específica do par O-D Zona Sul - Centro para as metodologias de k-fixa, k-adaptativo e os valores oficiais do PDTU. Observe que, ao considerar a linha de base do PDTU como a medida oficial de viagens, o fator k-fixa subestima os valores, situando-se abaixo da metade, ao passo que o k-adaptativo superestima-os, elevando as regiões do topo (dias úteis) para valores superiores ao dobro em relação à linha de referência.

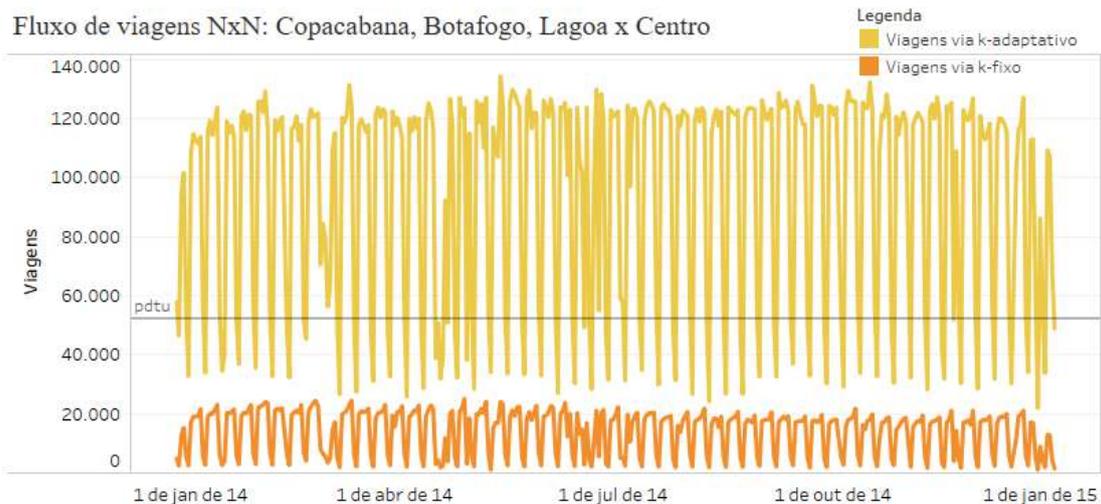


Figura 5.7: Quantidade de viagens zona sul x centro, visto através da matrizes NxN obtida via k-fixa e via k-adaptativo.

Para confirmar o padrão Casa-Trabalho, é preciso verificar o padrão de retorno. Ou seja, do Centro para Zona Sul (Trabalho-Casa). Esse padrão será apresentado a seguir.

Trabalho - Casa

Sob o ponto de vista de que o trabalho é um destino, ou um polo atrator, o padrão de mobilidade “Trabalho-Casa”(neste caso inferido por Centro x Zona Sul), pode ser compreendido como a volta para casa. A finalidade de verificar a volta é confirmar, através dos dados, se o padrão se mantém. As matrizes HBO e NHB facilitam essa verificação, pois sabe-se claramente se o indivíduo que originou a viagem reside ou não no local de partida.

O padrão regular Casa-Trabalho é visto novamente sob a direção Trabalho-Casa (Fig. 5.8). O padrão das viagens HBO e NHB revelam que a maioria das pessoas que saiu do Centro em direção à Zona Sul não reside no centro.

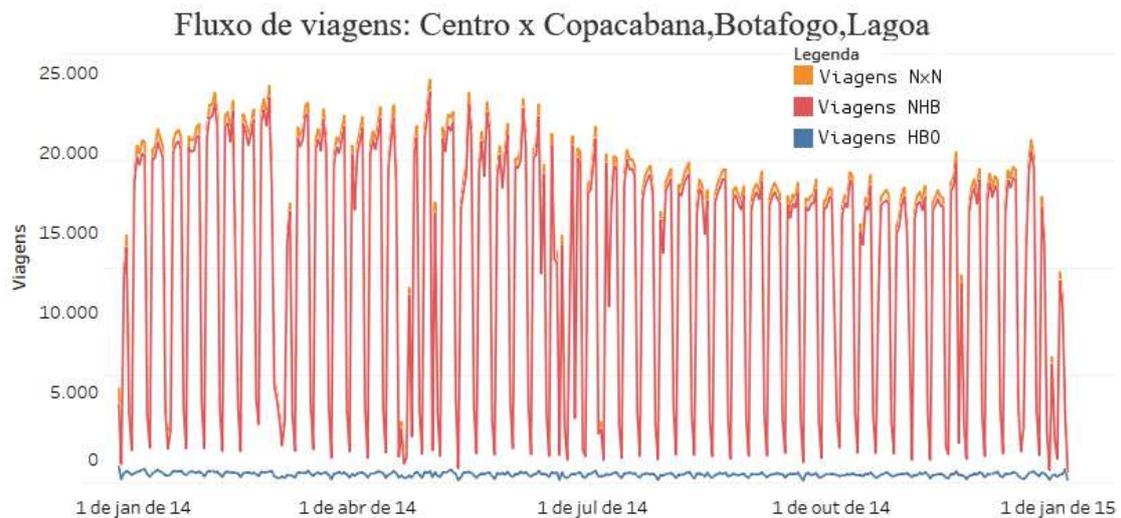


Figura 5.8: Quantidade de viagens Centro x Zona Sul (k-fixa), visto através das matrizes NxN, NHB e HBO.

Observa-se que o padrão regular de viagens NxN se manteve no percurso de volta, pode-se inferir também que esse padrão de mobilidade seja simétrico. Observando apenas o total de viagens (matriz NxN), é possível ver esse padrão simétrico durante todo o ano (Fig. 5.9), assim como sob as médias de viagens por dias da semana (Fig. 5.10). A simetria acompanha os feriados prolongados. Assim é possível confirmar o padrão de mobilidade com o objetivo de trabalho, dada a alta incidência de viagens nos dias úteis e baixíssima nos fins de semana, em ambos os sentidos (Fig. 5.10).

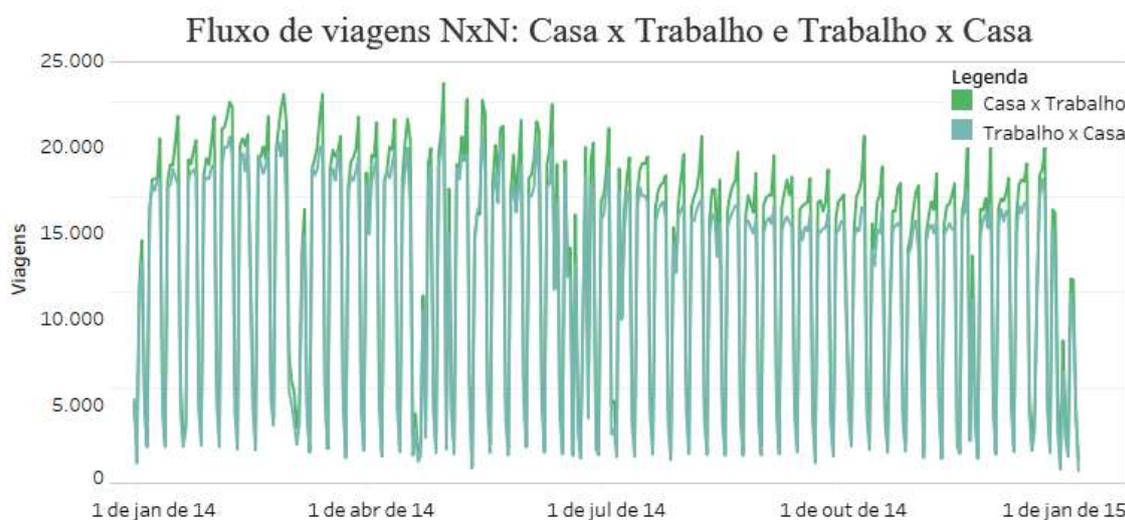


Figura 5.9: Quantidade de viagens Casa x Trabalho e Trabalho x Casa (k-fixa), equivalente a Centro x Zona Sul e Zona Sul x Centro, respectivamente, visto apenas pelas matrizes NxN.

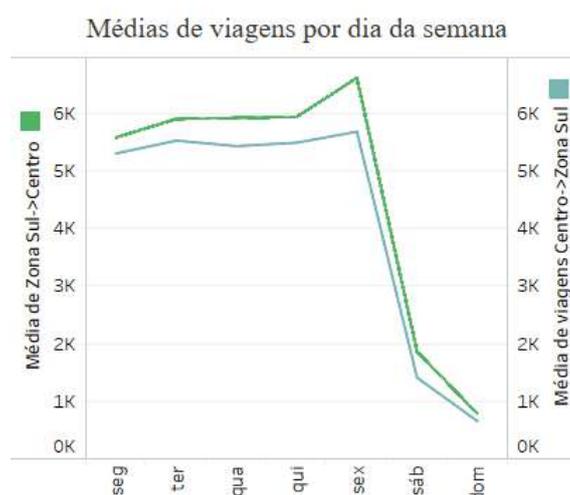


Figura 5.10: Média de viagens NxN (k-fixa) por dia da semana entre subdistritos da zona sul e centro; e vice-versa.

Casa - Lazer, Lazer - Casa

Em relação ao padrão Casa-Trabalho, o padrão Casa-Lazer é mais subjetivo e sujeito à preferências pessoais. O comportamento de mobilidade da população em feriados prolongados e fins de semana foi o principal critério para inferir o motivo de lazer, ou “caminho intermediário” ao lazer, nos destinos. Um critério adicional, foi observar o padrão de “volta” do destino, de onde fica claro que a maior parte das pessoas não voltou no mesmo dia, permanecendo um período

no destino e tornando o padrão assimétrico, motivo pelo qual os resultados serão apresentados com a “ida” e a “volta” no mesmo gráfico.

Segundo os critérios descritos acima, o trecho Zona Sul - Teresópolis foi o que apresentou um padrão mais definido. Pode-se observar que a maior quantidade de viagens ocorre no início de feriados, como a sexta-feira de carnaval em 28/04/2014 e a sexta-feira santa em 18/04/2014 (Fig. 5.11). Ao se observar a média semanal de viagens, os deslocamentos mais intensos da ida ocorrem nas sextas-feiras e nos sábados e a maioria dos deslocamentos da volta ocorre aos domingos (Fig. 5.12).

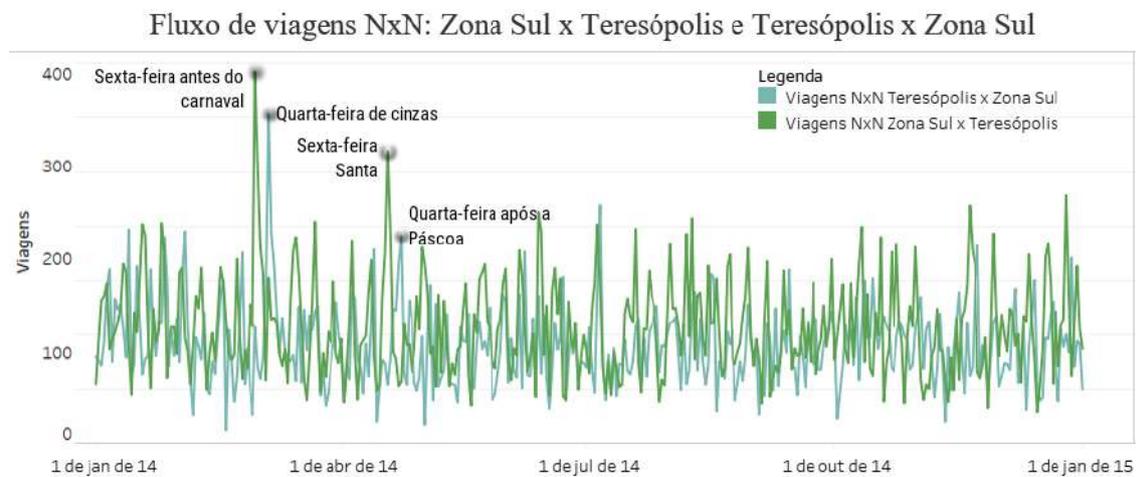


Figura 5.11: Quantidade de viagens Zona Sul x Teresópolis e Teresópolis x Zona Sul (k-fixa), visto pela matriz NxN.

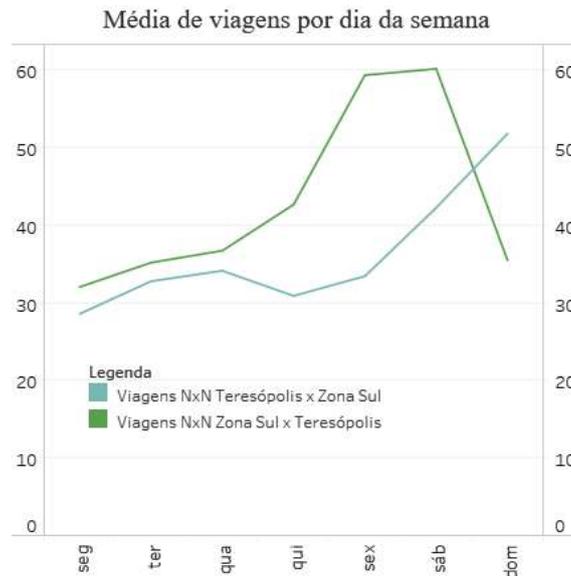


Figura 5.12: Média de viagens por dia da semana entre Zona Sul x Teresópolis e Teresópolis x Zona Sul (k-fixo), visto pela matriz NxN.

Sobre a quantidade de viagens entre o par O-D Zona Sul - Teresópolis, não há uma linha de base oficial, já que, Teresópolis não faz parte da RMRJ. Mesmo assim, é possível observar as diferenças entre o fator k-fixo e o k-adaptativo na Figura 5.13. Além do topo mudar de sexta-feira de Carnaval (k-fixo Fig.5.11) para Sexta-feira Santa (k-adaptativo), o valor do topo é 10x superior através do k-adaptativo, ou seja, de 400 para 4000 viagens.

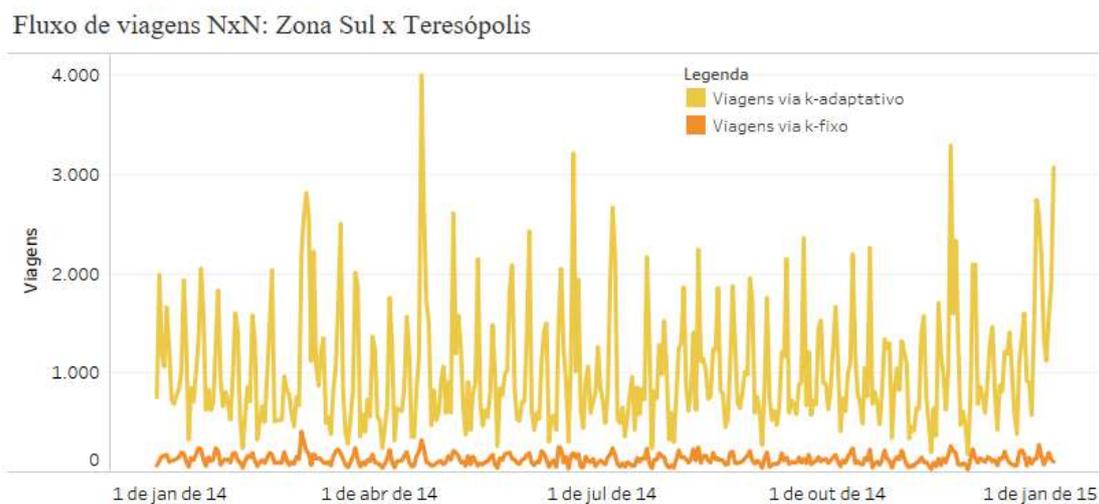


Figura 5.13: Quantidade de viagens Zona Sul x Teresópolis vista pela matriz NxN, obtida via k-fixo e via k-adaptativo.

A Região dos Lagos (Cabo Frio, Búzios, etc) é popularmente conhecida como destino de Lazer de grande parte da população da área de estudo. Apesar dessa região não estar contemplada nesta pesquisa, é possível abrir algumas discussões sobre os prováveis caminhos intermediários observados no *cordon line* (Fig. 5.14).

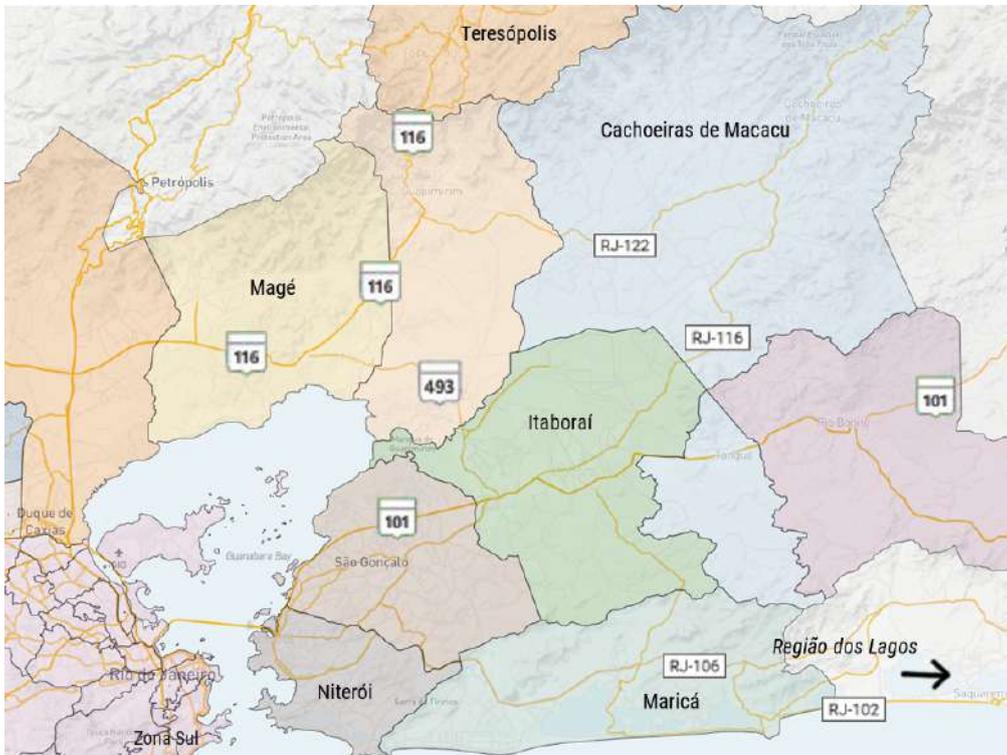


Figura 5.14: Principais eixos viários rumo à Teresópolis e Região dos Lagos.

Quem reside na cidade do Rio de Janeiro e costuma ir à Região dos Lagos, sabe que especialmente no Carnaval vai enfrentar longos engarrafamentos para sair da cidade². Os principais eixos viários implicam numa passagem por Niterói ou por Magé (ver mapa da Figura 5.14), via RJ-106/RJ-102 ou BR-116, respectivamente. A passagem por Niterói implica numa passagem por Maricá, antes de sair da região de estudo. Ao se observar as viagens NxN desse par O-D (Fig. 5.15), é visível o pico por conta do carnaval, e através das viagens NHB da mesma

²Esses engarrafamentos são largamente veiculados na mídia: O Globo <https://oglobo.globo.com/rio/engarramento-gigantesco-na-viagem-para-regiao-dos-lagos-11758147>, Extra <https://extra.globo.com/noticias/rio/saida-do-rio-tem-engarramento-para-regiao-dos-lagos-11755510.html> e portal r7 <https://noticias.r7.com/transito/noticias/volta-do-carnaval-motoristas-devem-evitar-pegar-a-estrada-a-partir-das-15h-desta-terca-feira-20140304.html>

O-D (Fig. 5.16) é possível observar que a maior parte das pessoas está apenas passando por Niterói.

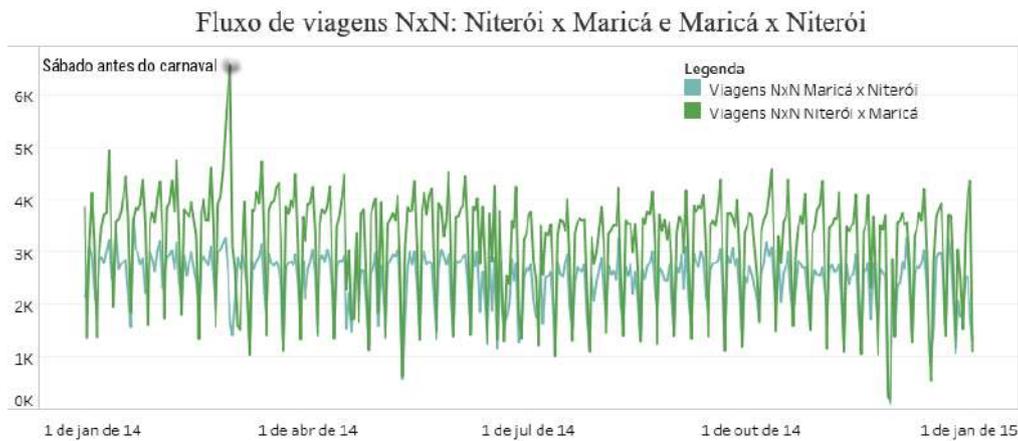


Figura 5.15: Viagens Niterói x Maricá e Maricá x Niterói, visto pela matriz NxN.

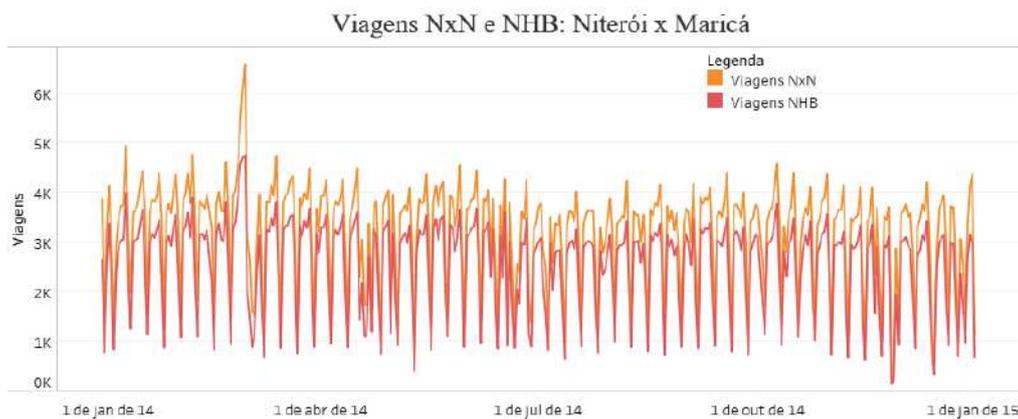


Figura 5.16: Viagens Niterói x Maricá, visto pelas matrizes NxN (laranja) e NHB (vermelho).

O trecho Niterói x Maricá pertence à RMRJ e conseqüentemente à abrangência do PDTU, de onde estima-se que ocorram 5.034 viagens em um dia útil comum. Esse valor ficou muito próximo do obtido a partir da quantidade de viagens via k-fixa, e ao mesmo tempo muito distante do obtido via k-adaptativo. Na Figura 5.17 é possível observar que os dois fatores de expansão foram consistentes em apontar o Sábado antes do Carnaval como um pico, registrando 54.028 viagens via k-adaptativo e 6.571 via k-fixa. O padrão de comportamento dos fatores de expansão observados em Niterói x Maricá foi similar ao caso a seguir, Magé x Itaboraí.

Fluxo de viagens NxN: Niterói x Maricá

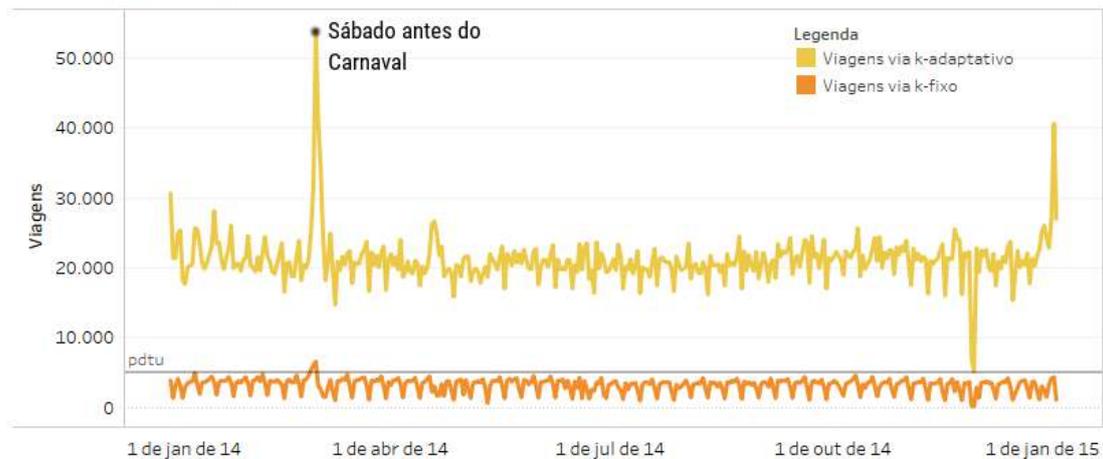


Figura 5.17: Quantidade de viagens Niterói x Maricá vista pela matriz NxN, obtida via k-fixo e via k-adaptativo.

O feriado do carnaval parece exercer bastante influência no comportamento de mobilidade da população. Ao se considerar um ponto de partida comum, como Magé via BR-116, foram identificados três trajetos com aumento de fluxo no carnaval: *a*) Magé x Itaboraí (Fig. 5.18); *b*) Magé x Cachoeiras de Macacu (Fig. 5.19); e *c*) Magé x Teresópolis (Fig. 5.20). Magé x Itaboraí expressa também a volta do feriado de Carnaval na Quarta-feira de Cinzas, embora esse comportamento não seja um padrão para todos os fins de semana. Um pico “somente de ida” pode ser visto no trajeto Magé x Cachoeiras de Macacu no sábado antes do Carnaval. Magé x Teresópolis aponta para dois picos, um antes do feriado do Carnaval e outro antes da Páscoa.

Padrões repetitivos são altamente previsíveis, nesse sentido torna-se mais fácil de prever em que dias quais vias ficarão congestionadas.

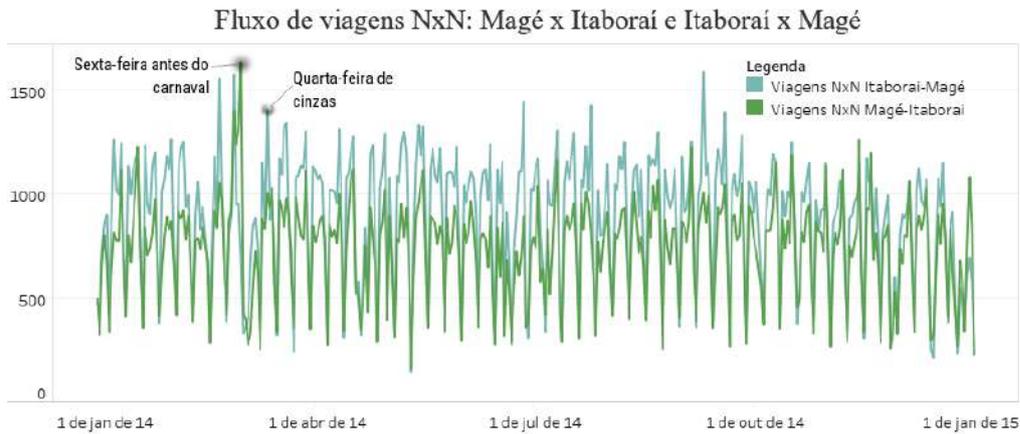


Figura 5.18: Quantidade de viagens Magé x Itaboraí e Itaboraí x Magé, visto pela matriz NxN.

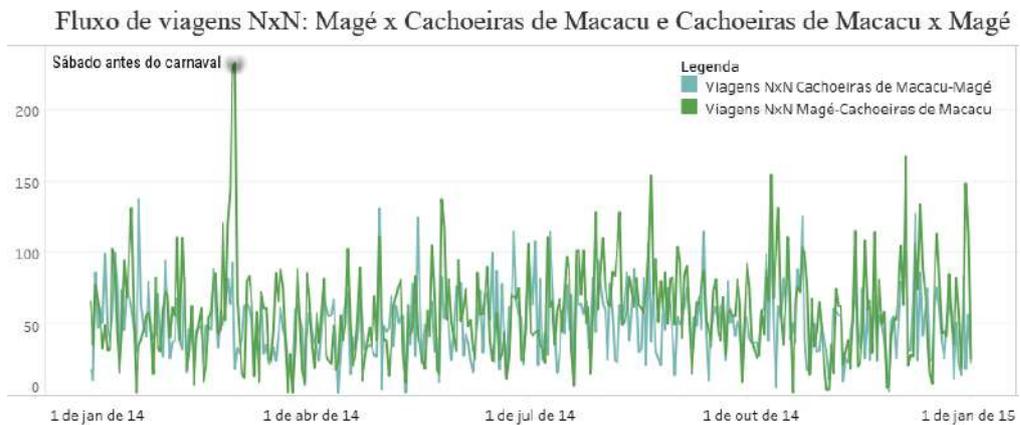


Figura 5.19: Quantidade de viagens Magé x Cachoeiras de Macacu e Cachoeiras de Macacu x Magé, visto pela matriz NxN.

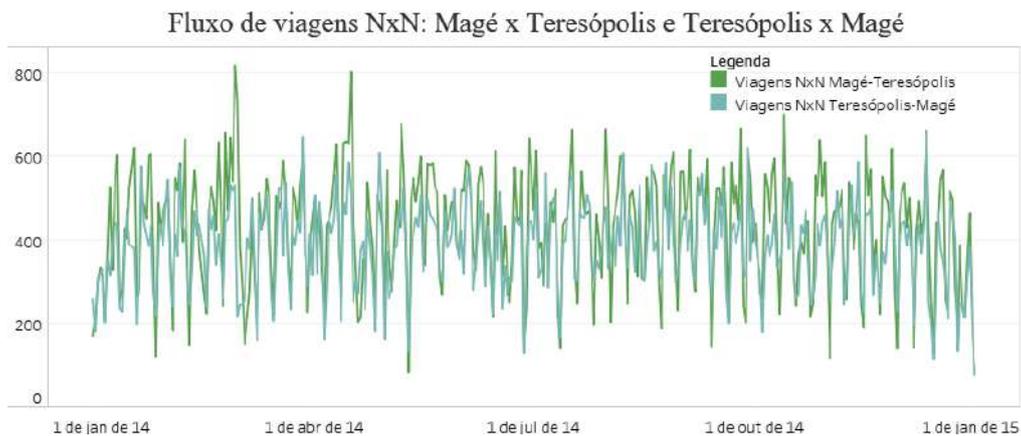


Figura 5.20: Quantidade de viagens Magé x Teresópolis e Teresópolis x Magé, visto pela matriz NxN.

5.5 Detecção de mudança em padrões de mobilidade

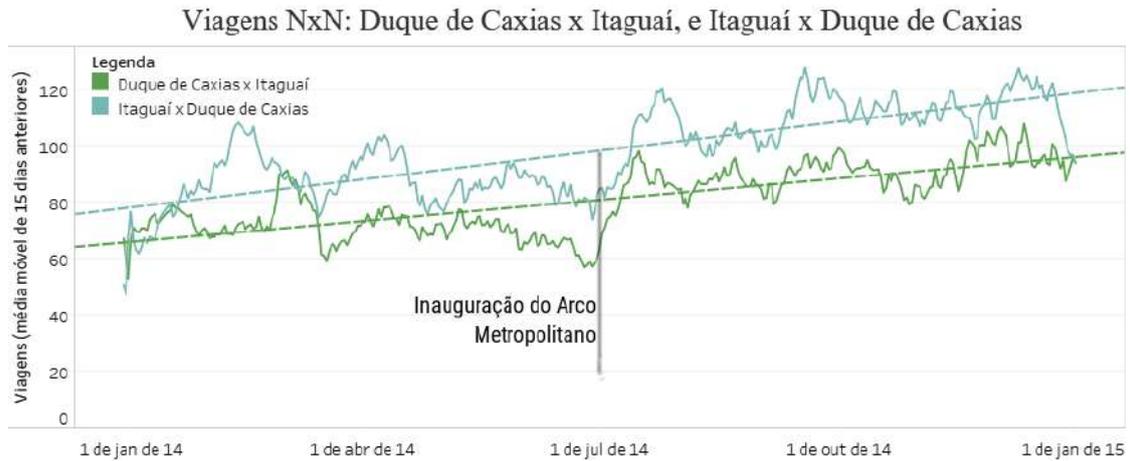
O primeiro critério adotado para considerar uma mudança de padrão foi observar através do padrão de viagens, quais percursos apresentaram uma mudança de comportamento que veio a tornar-se repetitiva. O segundo critério foi encontrar a data de inauguração de uma obra viária que fosse próxima da própria mudança.

Em geral, levantamentos como o do PDTU têm um custo muito elevado e são realizados uma vez a cada década, ou mais. A possibilidade de cálculo de uma matriz O-D por dia representa um grande potencial para a operação e monitoramento de uma grande cidade como o Rio de Janeiro. No ano de 2014 o Rio de Janeiro passou por diversas transformações viárias. Uma das alterações viárias importantes foi a inauguração da rodovia Rafael de Almeida Magalhães, conhecida como o Arco Metropolitano do Rio de Janeiro, em 1º de julho de 2014. O trecho inaugurado em 2014 liga a rodovia BR 101 (Rio – Santos) no município de Itaguaí com a rodovia BR 040 (Washington Luiz) em Duque de Caxias, conectando o trecho já existente da BR 493 até a BR 101 em Manilha. O Arco Metropolitano conecta as principais vias de acesso ao Rio de Janeiro, evitando que veículos precisem atravessar a cidade, reduzindo assim o tráfego nas vias expressas.

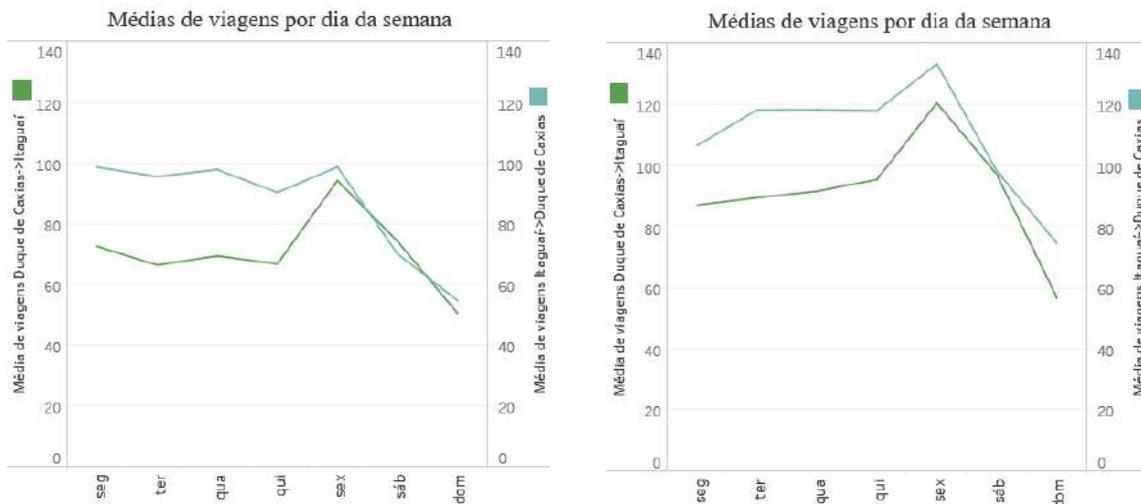
A mudança de comportamento provocada pela oferta da nova via (Arco Metropolitano) foi detectada nos deslocamentos diários entre os municípios de Duque de Caxias e Itaguaí, estimados pelos dados de CDR. A Figura 5.21 mostra os deslocamentos entre Duque de Caxias x Itaguaí e Itaguaí x Duque de Caxias (a), onde os dados foram filtrados pela média móvel de 15 dias anteriores para facilitar a visualização. Também é visto que a média diária de viagens (b) mudou após (c) inauguração do Arco. Pode-se observar que houve um aumento consistente do fluxo nos dois sentidos após a obra. Verifica-se também que o fluxo entre Itaguaí e Duque de Caxias é consistentemente maior que o fluxo entre Duque de Caxias e Itaguaí. Este resultado comprova o potencial da utilização de dados de

CDR para planejamento de transportes, bem como o monitoramento e operação de grandes cidades.

Para suavizar a visualização das séries temporais, foram adotadas médias móveis entre dias anteriores ou centralizada entre dias anteriores e posteriores, dependendo do que gerasse uma melhor visualização.



(a) Duque de Caxias x Itaguaí, Itaguaí x Duque de Caxias



(b) Antes do Arco

(c) Após inauguração do Arco

Figura 5.21: Viagens (k-fixa) entre Duque de Caxias e Itaguaí (e vice-versa), média móvel de 15 dias anteriores. A barra em (a) indica a data de inauguração do Arco Metropolitano. A média de viagens/dia da semana, (b) e (c), comprovam maior aumento em dias úteis.

A mudança de comportamento de mobilidade entre Duque de Caxias x Itaguaí também pode ser observada através do fator de expansão populacional adaptativo, que para este par O-D, foi mais aderente ao PDTU do que o fator fixo, conforme pode ser visto na Figura 5.22.

Fluxo de viagens NxN: Duque de Caxias x Itaguaí

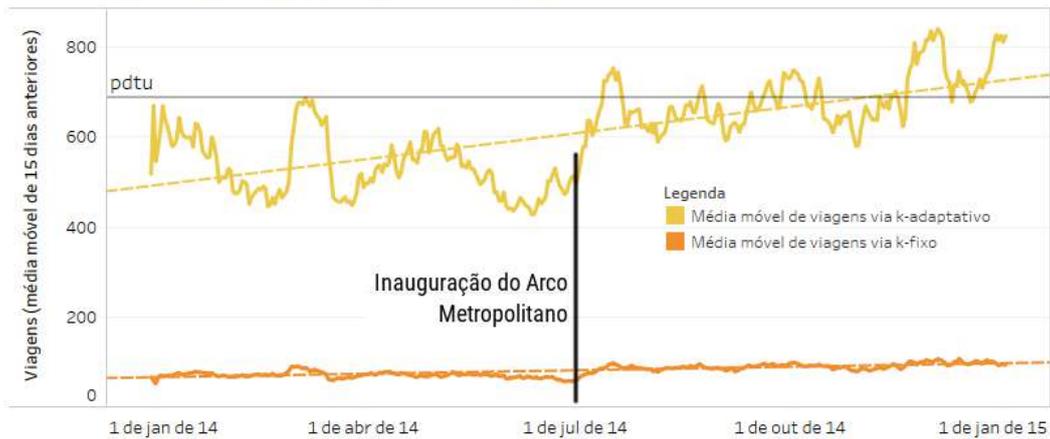


Figura 5.22: Quantidade de viagens Duque de Caxias x Itaguaí, visto pela matriz NxN obtida via k-fixa a via k-adaptativo.

Um mês antes da inauguração do Arco Metropolitano, houve a inauguração do BRT Transcarioca, em 1º de Junho de 2014³. Essa obra estabeleceu uma conexão viária entre a Barra (Terminal Alvorada) e o aeroporto do Galeão, passando por Madureira (Estação Madureira), Vicente de Carvalho, Olaria e Maré. Foi detectado um aumento do fluxo de viagens no trecho entre Barra e Madureira, (Fig. 5.23), contudo esse aumento não foi detectado na volta (Madureira x Barra). Observa-se que antes do BRT, a ida e a volta possuíam volumetrias relativamente proporcionais, contudo depois da inauguração, a quantidade de viagens aumentou apenas no fluxo Barra x Madureira. Esse mesmo comportamento foi observado sob o fator k-adaptativo.

³<http://www.rio.rj.gov.br/web/guest/exibeconteudo?id=5065490>, visto em junho de 2015.

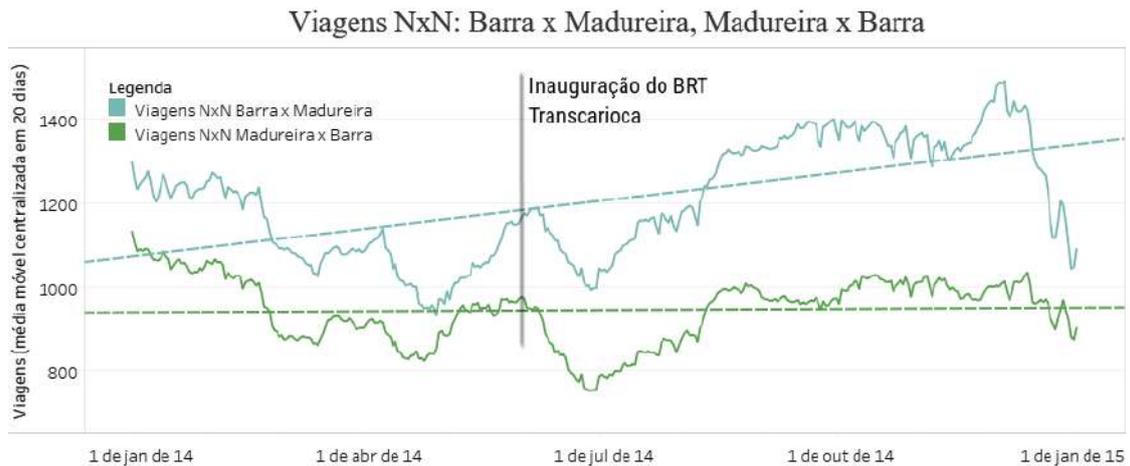


Figura 5.23: Média móvel de 20 dias (10 dias anteriores e 10 dias posteriores) das viagens NxN Barra x Madureira (k-fixa).

Um outro trecho pode ter mudado após a inauguração do BRT Transcarioca, trata-se da O-D Barra x Bangu, de onde se observa uma diminuição na quantidade de viagens (Fig. 5.24). Uma explicação possível dessa redução poderia ser atribuída ao simples fato de as pessoas terem mudado de trajeto para chegar em casa, beneficiando-se da conexão de trem entre Madureira e Bangu, contudo não foi observado aumento na volumetria de viagens desse outro par O-D.

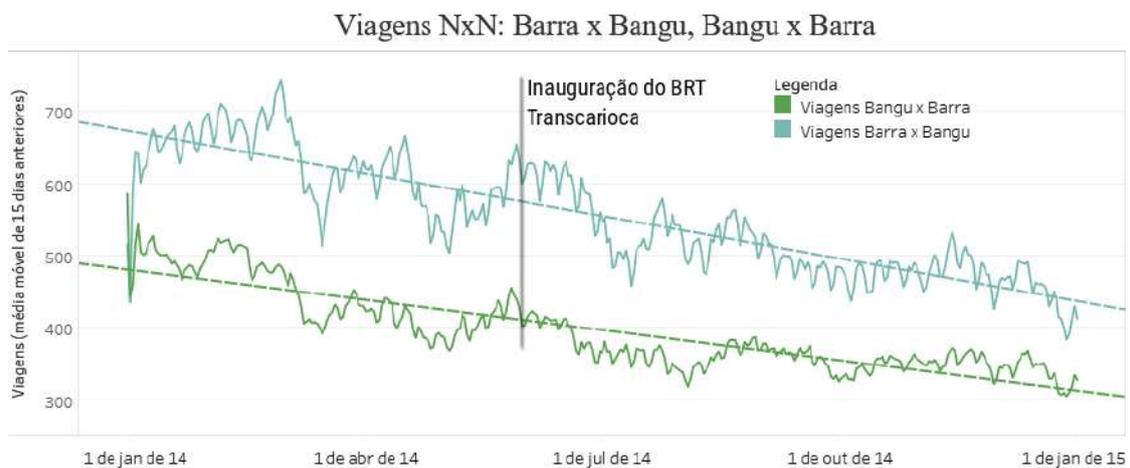


Figura 5.24: Média móvel de 15 dias anteriores das viagens NxN Barra x Bangu (k-fixa).

A diminuição progressiva de viagens no trecho Barra da Tijuca x Bangu também foi observada através do fator de expansão k-adaptativo, de onde se vê inclusive uma aderência muito próxima da linha de base do PDTU (Figura 5.25).

Fluxo de viagens NxN: Barra da Tijuca x Bangu

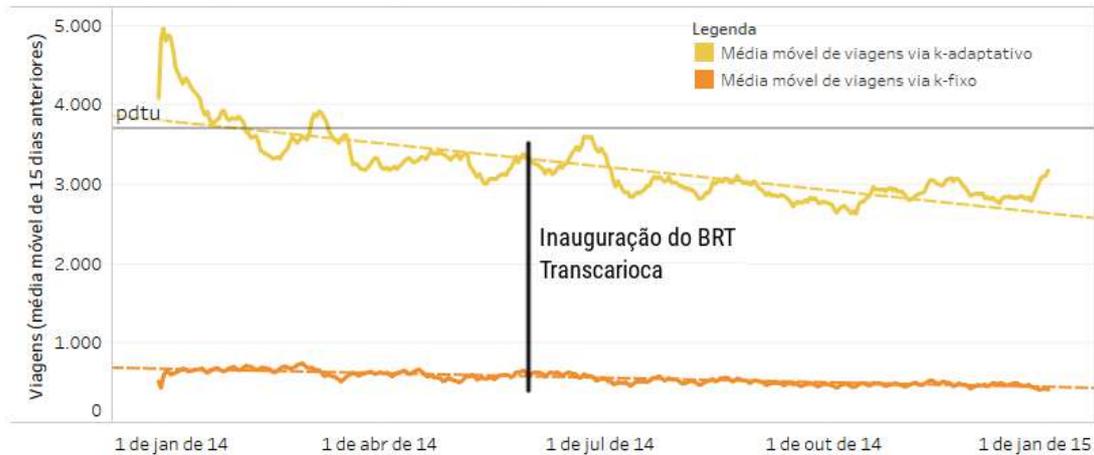


Figura 5.25: Quantidade de viagens NxN Barra x Bangu obtidas via k-fixa e via k-adaptativo.

Sobre os fatores de expansão populacional, observou-se que as linhas de base do PDTU, usadas como medidas de referência para a quantidade de viagens, oscilaram entre os valores obtidos via k-fixa e via k-adaptativo. Às vezes, as quantidades de viagens estimadas através dos dois fatores situavam-se sob escalas muito diferentes, dificultando a compreensão sobre qual o fator seria o mais próxima da realidade, pois a linha de referência ora se situa próxima do k-adaptativo, ora próxima do k-fixa. No capítulo de conclusão será proposto um direcionamento sobre esse dois fatores.

Foram mostradas as descobertas feitas a partir das 3025 (55x55) séries temporais obtidas a partir de todos os pares da matriz OD transiente, calculada segundo a metodologia explicada. Os padrões detectados encontraram suporte razoável na explicação de fenômenos já conhecidos de modo geral, que as pessoas seguem ciclos de mobilidade voltado principalmente a finalidades de trabalho e lazer. A possível relação de obras públicas com o comportamento de mobilidade da população tem potencial para ser uma ótima ferramenta para avaliar obras públicas, desde que a metodologia seja aprofundada com esse objetivo.

Capítulo 6

Resultados de matrizes O-D extraordinárias

*“Se a própria existência quotidiana
lhe parecer pobre, não a acuse.
Acuse a si mesmo, diga consigo
que não é bastante poeta para
extrair as suas riquezas.”*

— Rainer Maria Rilke, Cartas a um
jovem poeta.

Conteúdo

6.1	Validação da quantidade de pessoas em eventos	122
6.2	Visitantes nos blocos de carnaval em 2014	124
6.2.1	Descrição dos blocos	124
6.2.2	Estimativas segundo metodologia de CDR	126
6.3	Visitantes no Reveillon	129
6.3.1	Descrição dos locais	129
6.3.2	Estimativas segundo metodologia de CDR	132
6.4	Visitantes das praias do Rio de Janeiro durante o verão	135

O Rio de Janeiro é repleto de pontos de lazer dentro da própria cidade, há o Pão de Açúcar, o Cristo Redentor, o Jardim Botânico, as praias, apenas para citar alguns. A partir da granularidade de subdistrito, a tarefa de diferenciação

dos motivos de lazer, trabalho ou volta para casa, tornam-se desafiadoras, pois os locais de lazer são próximos ou até mesmo inseridos em áreas residenciais. Compreendendo esse contexto, viu-se a oportunidade de adotar delimitações específicas, para esclarecer padrões que ficaram diluídos dentro do esquema cotidiano. De forma geral, e na linha desta pesquisa, esses casos extraordinários são compreendidos como **eventos**.

O principal desafio ao se adotar uma delimitação espaço-temporal livre e específica para eventos, é que não há uma referência clara sobre a quantidade real de pessoas, e conseqüentemente viagens envolvidas. Os planos elaborados pela área de transporte geralmente não consideram a mobilidade por conta de eventos, de forma que esta pesquisa pretende colaborar nesse tema e propor uma abordagem de validação.

6.1 Validação da quantidade de pessoas em eventos

A contagem de pessoas em eventos geralmente é passada por autoridades públicas ou entidades reconhecidas, como é o caso da Datafolha ¹. Geralmente esses procedimentos remetem ao método proposto por Herbert Jacobs [72], onde é estimado o número de pessoas por metro quadrado e multiplicado pela área total do evento. Para estabelecer-se uma noção geral de pessoas por m^2 , pode-se considerar o PDTU 2013 [4], onde o limite de capacidade dos trens da SuperVia é medido considerando 8 pessoas/ m^2 (pág. 26).

A primeira dificuldade ao se inferir a quantidade de pessoas por m^2 em eventos ao ar livre, é que as pessoas geralmente estão aglomeradas de forma irregular; é preciso considerar um valor médio por m^2 que reflita a amplitude do evento. A segunda dificuldade é obter uma área delimitada “útil”, onde as pessoas podem ocupar um espaço físico, de fato. A terceira dificuldade é por conta de eventos

¹A Datafolha explica como estima a quantidade de pessoas em eventos, ver <http://www1.folha.uol.com.br/poder/2016/03/1756267-datafolha-tem-metodologia-para-calcular-multidao-variavel-entenda.shtml>, visto em 28 de Outubro de 2017.

itinerantes, onde a natureza do evento prevê que as pessoas deverão movimentar-se, em blocos, de um local a outro.

Para auxiliar na tarefa de contagem de pessoas em eventos, CHOI-FITZPATRICK e JUSKAUSKAS [73] desenvolveram uma metodologia a partir do uso de “*drones*” e software livre, onde uma dimensão de m^2 é relacionada a uma medida de $pixel^2$, dependendo da altura em que o *drone* tirar fotos da multidão. Essa técnica permite também acompanhar a durabilidade de um evento, pois a data e a hora da foto ficam registradas. Sob objetivos similares, NAINI *et al.* [74] desenvolveram uma técnica capaz de inferir a quantidade de pessoas em um determinado local através da detecção de celulares via tecnologia de Bluetooth². Uma outra alternativa seria inserir pessoas dentro do evento para acompanhar e contar as pessoas, como é o caso da Datafolha.

No caso dos resultados que serão apresentados a seguir, a presença das pessoas é estimada pela realização de no mínimo uma chamada telefônica na área de cobertura de um conjunto de torres de telefonia (ver exemplo na Figura 6.1), é preciso compreender a questão da área útil como uma limitação da metodologia, já que a área de cobertura das torres de telefonia abrange áreas sob o formato irregular dos polígonos de voronoi. Ao se escolher determinada torre, há o risco de cobrir-se uma região maior ou menor do que o desejado, essa limitação fica clara ao se observar a Figura 6.1, onde a área de cobertura da torres (roxo) é muito superior à área das ruas percorridas (amarelo). Assim sendo, e para futuras comparações, será mostrado um indicador de pessoas por m^2 levando em conta a área média das vias ou locais públicos utilizados, e a área de cobertura total das torres de telefonia consideradas no algoritmo.

²Trata-se de uma tecnologia de comunicação sem fio para curtas distâncias, disponível na maioria dos aparelhos celulares, <https://en.wikipedia.org/wiki/Bluetooth>, visto em 28/03/2018.

que 10 pessoas estavam compartilhando o mesmo m^2 ao mesmo tempo, mas sim que durante todo o tempo do evento, 10 pessoas passaram pelo mesmo metro quadrado.

Banda de Ipanema (1), (2) e (3)

O bloco Banda de Ipanema ocorreu três vezes: a primeira em (1) 15-fev-2014 entre 17h30 e 23h, a segunda (2) no dia 1º-mar-2014 entre 17h30 e 23h, e a terceira em (3) 4-mar-2014 entre 17h30 e 23h . Os eventos ocorreram sob o mesmo percurso, iniciando na Rua Gomes Carneiro, seguindo pela Avenida Vieira Souto, seguindo pela Rua Joana Angélica, e finalizando na Rua Visconde de Pirajá até a Praça General Osório. Para obter a área total das ruas foi considerada apenas a parte asfaltada, sob uma largura ⁴ de 10m. A soma da área das ruas envolvidas é de $28810m^2$. Sendo que o público estimado foi de 80.000 pessoas por evento, temos uma estimativa de 2,77 pessoas/ m^2 a cada evento.

Simpatia é Quase Amor (1) e (2)

O bloco Simpatia é Quase Amor ocorreu duas vezes: a primeira em (1) 22-fev-2014 entre 16h e 21h, e a segunda em (2) 2-mar-2014 entre 16h e 21h. O percurso foi o mesmo para os dois, iniciando na Rua Teixeira de Melo (esq. com R. Visconde de Pirajá), e seguindo pela Avenida Vieira Souto (ocupando as duas pistas) até a Anibal de Mendonça. A região deste bloco é a mesma do bloco anterior, de forma que foi adotada a mesma medida padrão para largura de rua, 10m. A área total das ruas envolvidas é de $28470m^2$. Com um público estimado de 120.000 pessoas, temos uma estimativa de 4,21 pessoas/ m^2 por evento.

Bloco da Preta

O Bloco da Preta ocorreu no dia 23-fev-2014 entre 15h e 20h. O local foi a Av. Rio Branco, da Presidente Vargas até a Cinelândia. Dado que trata-se de uma

⁴Observada no Google Maps em 2/11/2017. Para o caso da Av. Vieira Souto, foram consideradas as duas pistas.

região mais antiga do Rio de Janeiro, com calçadas mais largas, foi considerada a largura de 30m, incluindo as calçadas. Assim sendo, a área total considerada é de $36300m^2$. Com um público estimado de 300.000 pessoas, temos uma estimativa de 8,26 pessoas/ m^2 .

Cordão do Bola preta

O bloco Cordão do Bola Preta ocorreu no dia 1º-mar-2014 entre 9h e 16h. O local foi o mesmo do evento anterior. Com um público estimado de 1.500.000, temos uma estimativa de 41,32 pessoas/ m^2 .

Monobloco

O bloco Monobloco ocorreu no dia 9-mar-2014 entre 9h e 16h. O local foi o mesmo dos dois blocos anteriores. Com um público estimado de 500.000, temos uma estimativa de 13,77 pessoas m^2 .

Sargento Pimenta

O bloco Sargento Pimenta ocorreu no dia 3-mar-2014 entre 15h e 20h. Esse bloco tem uma grande diferença em relação aos demais, pois ele não percorre nenhuma avenida, fica parado no Aterro do Flamengo, próximo ao Monumento dos Pra-cinhas, não havendo referencia alguma sobre a área que ocupou. O público estimado foi de 100.000 pessoas.

6.2.2 Estimativas segundo metodologia de CDR

Os visitantes de blocos de carnaval foram contados a partir da realização de no mínimo uma chamada telefônica nos locais e intervalos de tempo descritos na seção anterior. A partir das áreas públicas das vias utilizadas e das áreas de cobertura das antenas, foi montado um indicador de pessoas/ m^2 . Os resultados estão detalhados na Tabela 6.1. O indicador de pessoas/ m^2 considerado através das áreas das torres de telefonia não mostrou relação com o das áreas úteis. Isso

pode ser visto na Figura 6.2. Uma possível explicação para essa falta de relação é o formato irregular inerente da cobertura das antenas.

Bloco de Carnaval	Visitantes (segundo CDR)	Pessoas/ m^2 (área útil)	Pessoas/ m^2 (área das torres)
Bloco da Preta	183.128	5,04	0,027
Simpatia é quase amor (1)	37.337	1,31	0,055
Simpatia é quase amor (2)	87.784	3,08	0,128
Sargento Pimenta	132.631	-	0,062
Cordão do Bola Preta	286.132	7,88	0,042
Monobloco	145.796	4,04	0,022
Banda de Ipanema (1)	44.175	1,53	0,053
Banda de Ipanema (2)	93.950	3,26	0,113
Banda de Ipanema (3)	73.588	2,55	0,088

Tabela 6.1: Visitantes nos blocos de carnaval do Rio de Janeiro em 2014 segundo estimativas da metodologia de CDR.

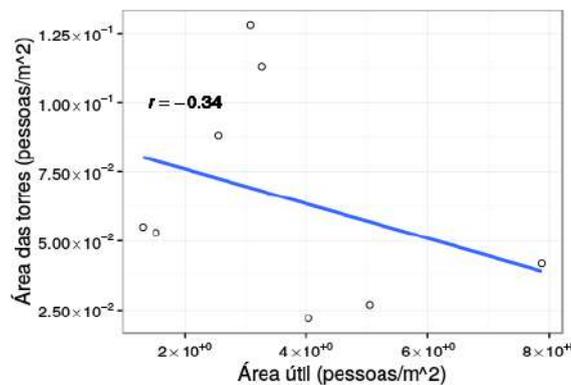


Figura 6.2: Aparente falta de relação entre pessoas/ m^2 sob os dois tipos de área: útil e torres de telefonia.

O local inferido como origem dos visitantes é a residência presumida, pois os resultados foram obtidos a partir de apenas uma chamada. A volumetria da população visitante por local de origem pode ser vista na Figura 6.3 sob o enfoque nas regiões da RMRJ que mais produziram viagens aos blocos. O posicionamento dos gráficos no mapa está ligado às coordenadas geográficas das regiões de residência. O diâmetro dos gráficos de setores está ligado a quantidade de pessoas que estava em algum dos blocos. Os setores evidenciam os blocos escolhidos, de acordo com as cores. Observe que para os blocos que se repetiram em vários dias⁵ não foi feita nenhuma agregação, considerando a mobilidade de cada evento como individual.

⁵Banda de Ipanema e Simpatia é Quase Amor.

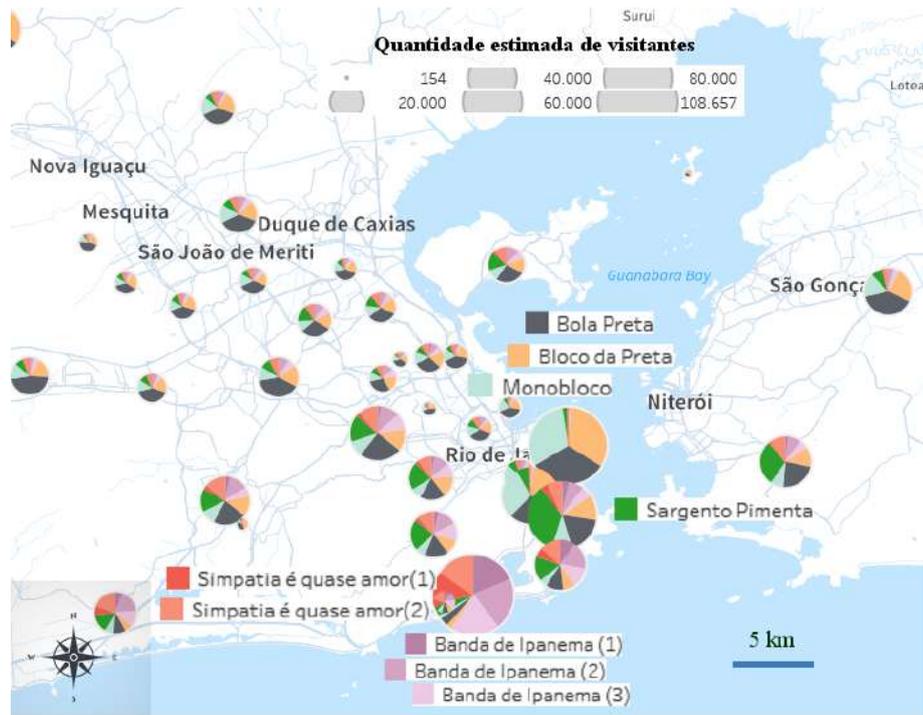


Figura 6.3: Quantidade de visitantes aos principais blocos de carnaval de 2014 na cidade do Rio de Janeiro.

A probabilidade dos indivíduos escolherem um determinado bloco em função da distância a percorrer ⁶ pode ser vista na Figura 6.4. Conforme visto na metodologia, a probabilidade é obtida pela razão entre os indivíduos que realizaram alguma chamada no bloco X e todos os usuários que realizaram alguma chamada no intervalo dos blocos listados no início desta seção, independente do local. Cada ponto é um local de origem a cada cor é um bloco, de forma que existem tantos pontos quanto locais de origem x blocos. Os pontos no topo evidenciam os blocos cuja probabilidade de receber visitantes é maior do que os de baixo. É de onde se reafirma a preferência pelo Cordão do Bola Preta, representado pelos pontos pretos. Na sequência das preferências, aparecem os pontos laranja e azul claro, indicando Bloco da Preta e Monobloco respectivamente. A Banda de Ipanema (em tons de roxo para os três eventos), o Simpatia é Quase Amor (em tons de vermelho para os dois eventos) e o Sargento Pimenta, apresentam um padrão mais difuso.

⁶Distância através das vias de acesso, obtida através da *API Distance Matrix* do Google Maps.

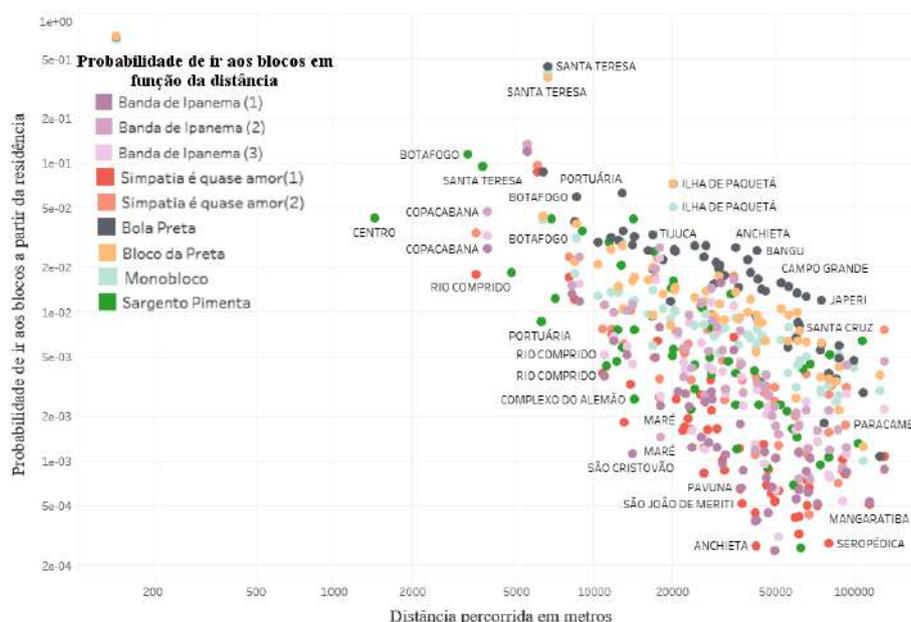


Figura 6.4: Probabilidade de um indivíduo ir um bloco de carnaval vs. a distância a percorrer a partir da própria residência.

6.3 Visitantes no Reveillon

O reveillon é tipicamente um dos eventos mais celebrados no mundo. Geralmente os grandes centros urbanos contam com eventos compostos de shows e queima de fogos que atraem um grande público. Os principais locais da RMRJ que anunciaram queima de fogos para o reveillon de 2014 serão descritos a seguir. Para os locais em que houve divulgação, será mostrado o público estimado pela prefeitura ou entidades responsáveis, após a ocorrência do evento ⁷.

6.3.1 Descrição dos locais

Os locais escolhidos da RMRJ estão distribuídos em três cidades: *a*) Rio de Janeiro (em Copacabana, Flamengo, Madureira e Penha), *b*) Niterói (na Praia de Icaraí), e *c*) Duque de Caxias (na Praça do Pacificador).

⁷Fonte: <http://www.rio.rj.gov.br/web/guest/exibeconteudo?id=4530645>, visto em 3/11/17

Copacabana

Em 2014 a praia de Copacabana estava com infraestrutura de três palcos dispostos na orla⁸, cada um deles com diferentes atrações ocorrendo de forma simultânea. Segundo medição de área feita através de imagem de satélite do Google Maps⁹, a área total da orla de Copacabana é de $448.000m^2$. A medição considerou toda a extensão de areia da orla, finalizando na calçada e iniciando onde a areia estivesse seca. É possível que se encontre uma dimensão diferente se as imagens de satélite forem atualizadas em maré baixa ou alta. Segundo noticiado na mídia¹⁰, 2 milhões de pessoas assistiram a queima de fogos em Copacabana, tanto em 2014 quanto em 2015. Considerando a quantidade de pessoas por área útil, temos uma estimativa de 4,46 pessoas/ m^2 .

Flamengo

A praia do Flamengo usualmente também conta com atrações de shows além da queima de fogos. A área da orla da praia do Flamengo é de $93500m^2$, desde a parte da areia seca até o final do calçadão. A estimativa da prefeitura é que 300.000 pessoas estiveram nesse evento. Considerando a área útil, temos uma estimativa de 3,21 pessoas/ m^2 .

Madureira

O Parque Madureira possui uma área útil bem definida, $93.553m^2$, pois foi inaugurado em 2012 e recebeu ampliações após o ano de 2015. O projeto do Parque Madureira pode ser visto com mais detalhes na dissertação de mestrado do secretário de obras da prefeitura do Rio de Janeiro, Mauro Chagas Bonelli [75].

⁸Ver <http://g1.globo.com/rio-de-janeiro/noticia/2013/12/reveillon-no-rio-tera-shows-gratis-em-9-bairros-queima-de-fogos-e-beijaco.html>, visto em 3/11/2017.

⁹Visto em 3/11/2017

¹⁰Fonte:<http://g1.globo.com/rio-de-janeiro/noticia/2014/01/reveillon-do-rio-emociona-com-fogos-tematicos-beijaco-e-calor.html> e <http://agenciabrasil.ebc.com.br/geral/noticia/2016-01/queima-de-fogos-no-reveillon-de-copacabana-e-vista-por-2-milhoes-de-pessoas>, visto em 3/11/2017.

Segundo estimativas da prefeitura, 40 mil pessoas foram ao Reveillon no Parque Madureira. Sendo que os eventos de Reveillon em estudo ocorreram antes das ampliações, a área do projeto inicial foi considerada como parâmetro. Assim sendo, temos uma estimativa de 0,43 pessoas / m^2 .

Penha

Segundo a prefeitura, a queima de fogos na Penha ocorreu no IAPI da Penha. Haja visto que trata-se de um conjunto habitacional com uma praça no centro, foi considerada a área dessa praça. Trata-se da praça Santa Emilianiana com uma área de $17.800m^2$, essa área leva em conta a extensão da rua Santa Engrácia até a rua Santa Basilissa. A prefeitura estimou que 20.000 pessoas assistiram a queima de fogos no IAPI da penha. Assim sendo, temos a estimativa de 0,89 pessoas / m^2 .

Niterói

A Praia de Icaraí possui uma área de $84.800m^2$. Foi considerada a mesma forma de medição da Praia de Copacabana e da Praia do Flamengo, ou seja, da parte seca da areia até o fim do calçadão. Segundo conteúdo noticioso, 400.000 pessoas assistiram à queima de fogos na Praia de Icaraí¹¹. Assim sendo, temos a estimativa de 4,71 pessoas/ m^2 .

Duque de Caxias

O palco dos eventos públicos em Duque de Caxias é a Praça do Pacificador, que possui uma área aproximada de $2100m^2$, desconsiderando a área do Teatro Raul Cortez e da Biblioteca Municipal Leonel Brizola. Não foi encontrada estimativa de pessoas que assistiram ao Reveillon em 2014 ou 2015 no evento.

¹¹Fonte: <http://soumaisniteroi.com.br/show-de-luz-e-cores-na-festa-de-reveillon-em-niteroi/>, visto em 3/1/17.

6.3.2 Estimativas segundo metodologia de CDR

Da mesma forma que os blocos de carnaval, a presença das pessoas nos eventos de Reveillon foi inferida através de uma chamada telefônica. Também foi considerado o esquema de área útil e de área de cobertura das torres de telefonia móvel. Na Figura 6.5 pode ser visto através do exemplo da Praia de Copacabana, como foram auferidas as áreas de todos os locais.



Figura 6.5: Mapa das áreas consideradas em Copacabana. A Orla (incluindo o calçadão) está destacada em amarelo, e a área de cobertura das torres de telefonia está destacada pelo polígono azul.

Os resultados de visitantes para o evento de Reveillon obtidos para o ano de 2014 e 2015 estão detalhados nas tabelas 6.2 e 6.3 respectivamente, conforme a seguir.

Local de Reveillon 2014	Visitantes (CDR)	Pessoas/m ² (área útil)	Pessoas/m ² (área das torres)
Praia de Copacabana	422.006	0,94	0,08
Praia do Flamengo	69.902	0,74	0,04
Penha (IAPI)	40.910	2,3	0,02
Praia de Icaraí (Niterói)	104.804	1,24	0,09
Praça do Pacificador (Duque de Caxias)	57.693	5,39	0,02
Parque Madureira	125.139	1,34	0,01

Tabela 6.2: Visitantes nos locais típicos de Reveillon na RMRJ em 2014, segundo estimativas da metodologia de CDR.

Local de Reveillon 2015	Visitantes (CDR)	Pessoas/m ² (área útil)	Pessoas/m ² (área das torres)
Praia de Copacabana	455.753	1,02	0,09
Praia do Flamengo	71.507	0,76	0,04
Penha (IAPI)	54.546	3,06	0,02
Praia de Icaraí (Niterói)	127.150	1,5	0,11
Praça do Pacificador (Duque de Caxias)	55.310	5,17	0,02
Parque Madureira	130.310	1,39	0,01

Tabela 6.3: Visitantes nos locais típicos de Reveillon na RMRJ em 2015, segundo estimativas da metodologia de CDR.

Os “visitantes” aos eventos de Reveillon foram contados apenas se realizaram no mínimo uma chamada entre às 21h de 31 de Dezembro de 2013 (ou 2014) e às 3h de 1º de Janeiro de 2014 (ou 2015) em alguma das regiões elencadas na seção anterior. A volumetria inferida da população visitante no ano de 2014 pode ser vista na Figura 6.6, que dá enfoque às regiões que mais produziram viagens a algum dos eventos de Reveillon. O diâmetro dos gráficos de setores (Fig 6.6) está ligado a quantidade de pessoas que estava em algum dos eventos. Os setores dos gráficos evidenciam o destino escolhido, de acordo com as cores. O posicionamento dos gráficos está ligado às coordenadas geográficas (latitude e longitude) da região de residência.

Um olhar atento na Figura 6.6 pode revelar que as pessoas tendem a passar o Reveillon próximo de casa. Para aguçar essa possibilidade, foi montada a Figura 6.7 com ativação de filtros para os principais eventos: (a) Praia de Copacabana, (b) Praia do Flamengo, (c) Praia de Icaraí, e (d) Parque Madureira. É possível observar que Copacabana é um destino que atrai uma boa parcela de residentes de diversas regiões. A queima de fogos no Flamengo atrai cerca da metade dos residentes do subdistrito Botafogo, que inclui o local onde houve a queima de fogos. Opondo-se a Copacabana, a queima de fogos em Niterói foi o maior polo atrator “do outro lado da ponte”. Na região da baixada, o local preferido foi o Parque Madureira. Em todos esses casos, observa-se uma tendência de as pessoas ficarem mais próximas de casa.

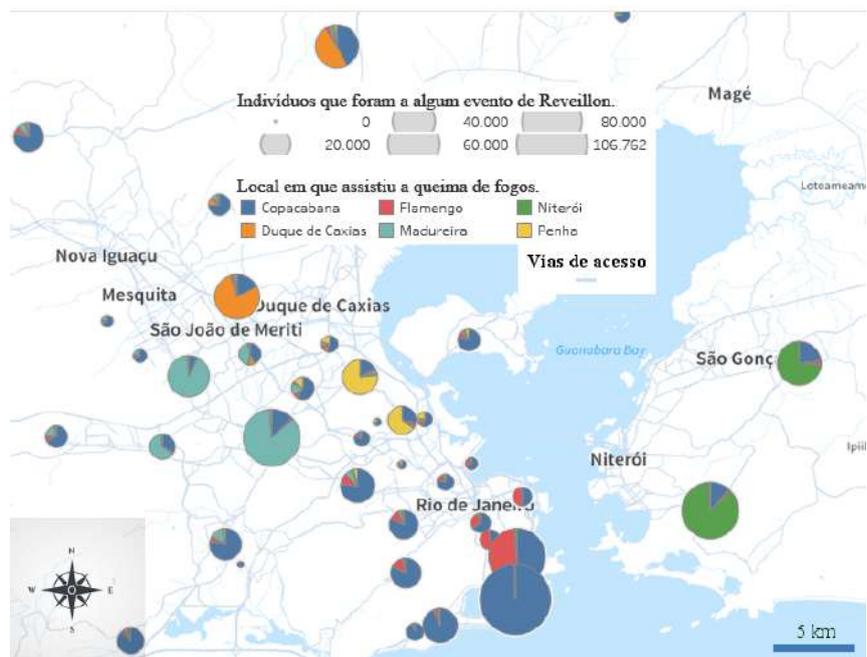


Figura 6.6: Distribuição geográfica das pessoas que foram a algum evento de Reveillon em 2014 onde houve queima de fogos.

A probabilidade dos indivíduos irem a um evento em função da distância a percorrer pode ser vista na Figura 6.8. Conforme visto na metodologia, a probabilidade é obtida através da razão entre a população que estava no evento pelo total de pessoas detectadas no intervalo de tempo do evento, independente de onde estavam. Para elucidar melhor a visualização da figura, a escala dos eixos é logarítmica e cada ponto é um local de origem, sendo a cor do ponto, o evento escolhido. Ou seja, há tantos pontos quanto locais x eventos. Quanto mais os pontos estão acima, maior a probabilidade, quanto mais estão à direita, maior a distância a percorrer a partir da própria região de residência até o evento. Assistir aos fogos em Copacabana tem alta probabilidade de ser a escolha da maioria dos residentes da área 21 (RMRJ), mesmo que se tenha que percorrer longas distâncias para isso. Também pode-se verificar uma tendência de as pessoas passarem o Reveillon perto de casa, é isso que indicam os dois primeiros pontos no alto a esquerda, a probabilidade dos residentes de Copacabana e de Madureira ficarem próximos de suas residências durante o evento.

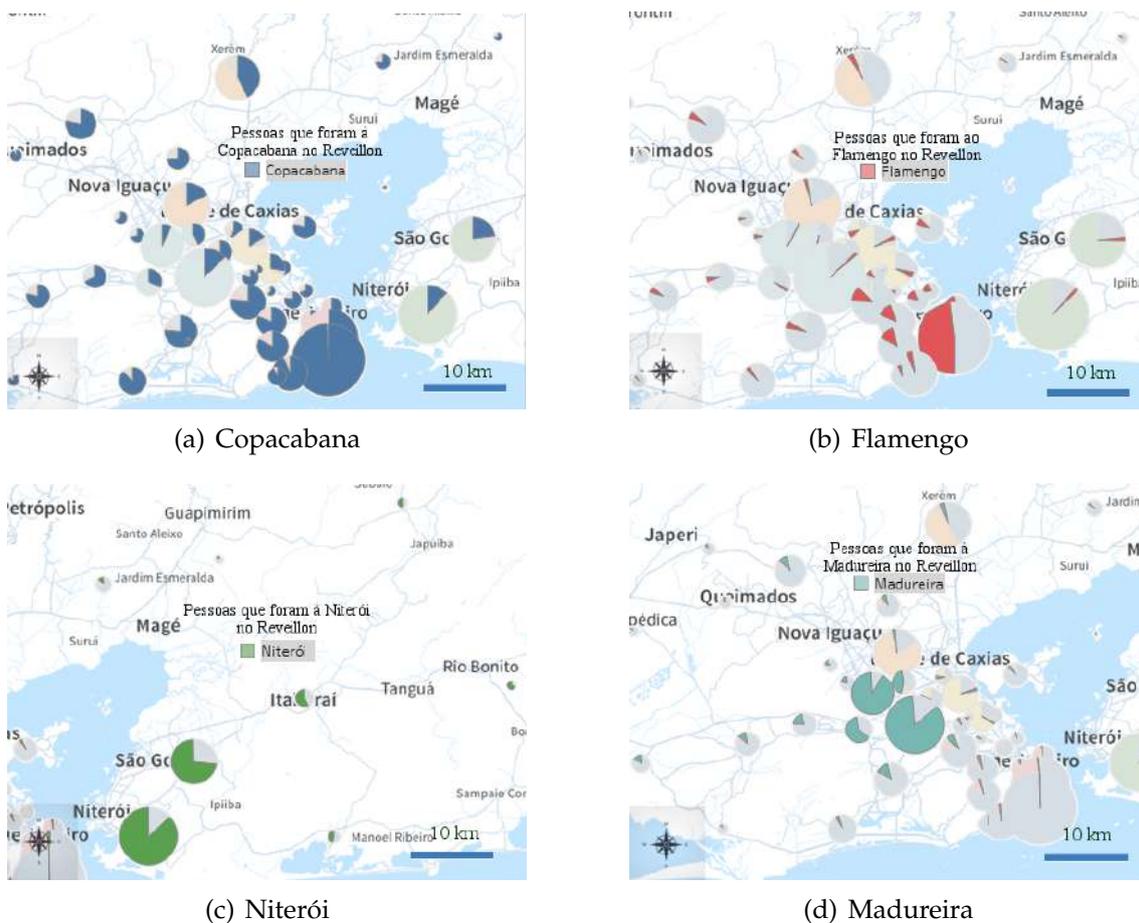


Figura 6.7: Volume de visitantes do Reveillon 2014, indicando a tendência de preferência por atrações próximas de casa.

6.4 Visitantes das praias do Rio de Janeiro durante o verão

Os dias de sol escaldante do verão carioca estimulam a população às praias. Considerando-se a vasta orla da cidade do Rio de Janeiro como um extenso polo atrator durante todos os domingos do verão de 2013/2014, foi possível observar as praias preferenciais dos moradores. Ao que consta, e até o presente momento¹², não há censo ou pesquisa que leve em conta a quantidade de pessoas nas Praias, de forma que não há uma base para estimativas ou validações.

Da mesma forma que nos resultados anteriores, foram consideradas duas áreas para apresentar a quantidade de pessoas/ m^2 : a) a área da orla em si, con-

¹²Até 5/11/2017.

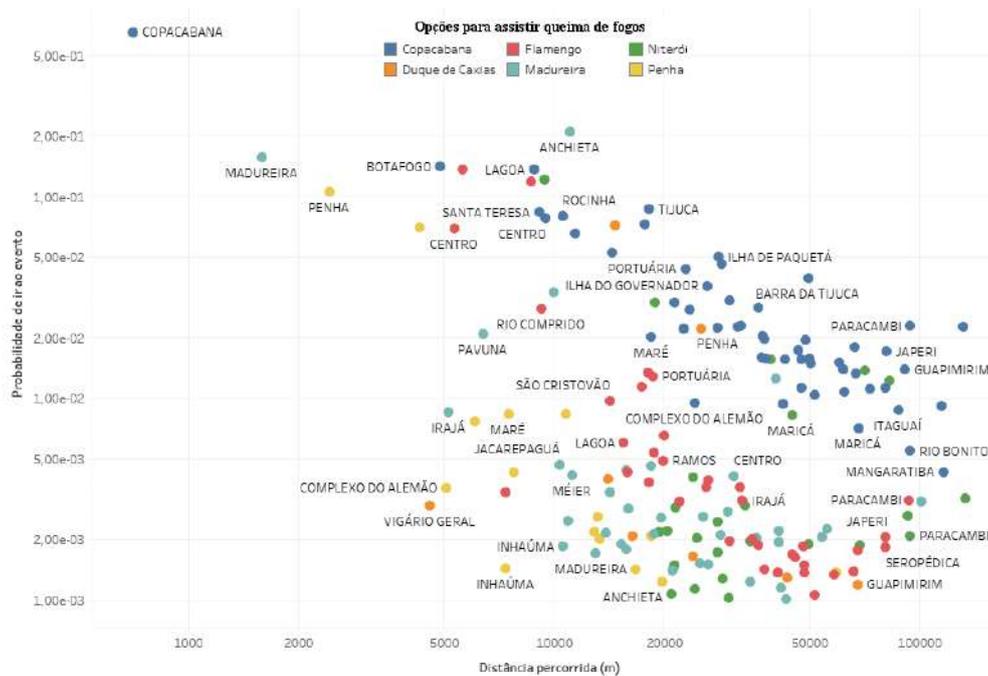


Figura 6.8: Probabilidade versus distância percorrida de indivíduos de determinado local irem aos destinos de Reveillon. A escala é logarítmica.

templando a extensão da areia seca vista na imagem do satélite¹³ até a pista fechada para lazer¹⁴, e *b*) a área do conjunto de antenas (*em polígonos de voronoi*) considerada pelo algoritmo.

Sob uma visão ampla da área de estudo (Fig. 6.9), é possível observar que a área de cobertura das torres de telefonia em algumas regiões vai muito além da orla, como é o caso da Praia da Reserva, localizada numa reserva ambiental, que possui uma área média de cobertura por torre de 240,55 hectares por antena (ver Tabela 6.4), diminuindo a precisão geográfica. Trata-se de um valor muito alto se comparado com áreas residenciais, como é o caso da média de 7,2 ha da Praia de Copacabana, por exemplo.

As praias consideradas foram: Leme, Copacabana, Arpoador, Ipanema, Leblon, São Conrado, Barra (dividida em Pepê, meio da Barra, Reserva, Recreio e Pontal). As áreas da orla e da cobertura das torres de telefonia podem ser vistas na Tabela 6.4, juntamente com a área de cobertura média por antena. Esse

¹³Fonte: Google Maps, visto em 2/nov/2017, ver o mapa publicado em : <https://tinyurl.com/ybabrehd> (necessário fazer *log-in* no Google).

¹⁴Considerando os locais em que as pistas fecham aos domingos: Leme, Copacabana, Arpoador, Ipanema, Leblon.



Figura 6.9: Praias da orla do Rio de Janeiro (do Leme ao Pontal) delimitadas pela cobertura dos polígonos de voronoi das torres de telefonia correspondentes.

último indicador foi adicionado a título de melhor contextualização e senso de proporção, já que as áreas de cobertura são muito distintas entre si e não foram encontradas bases para estimativas.

Praias (do Leme ao Pontal)	Área (ha) (orla)	Área (ha) \sum (torres telefonia)	Área (ha) média/torre
Praia do Leme	22,2	160	14,54
Praia de Copacabana	26	144	7,2
Praia do Arpoador	7,71	55	9,16
Praia de Ipanema	18,4	60	10
Praia do Leblon	13,9	72	18
Praia de São Conrado	15	126	42
Praia do Pepê (Barra)	19,5	207	34,5
Praia do Meio da Barra	16,5	243	48,6
Praia da Reserva	33	2165	240,55
Praia do Recreio	11,6	304	76
Praia do Pontal	7,96	830	276,67

Tabela 6.4: Principais praias do Rio de Janeiro com suas respectivas áreas de orla, cobertura total das torres de telefonia e cobertura média por torre.

A detecção da população nas praias é inferida pela realização de no mínimo uma chamada, processada por alguma das antenas que atendem às regiões das orlas. Foram contados como visitantes aqueles detectados entre 6h e 18h de todos

os domingos do verão em alguma das regiões destacadas na Figura 6.9. Os valores apresentados na Tabela 6.5 levam em conta a quantidade média de visitantes entre todos os domingos. Observe que entre as duas praias mais visitadas, Reserva e Copacabana, apesar da Reserva ter uma quantidade maior de visitantes, Copacabana tem 20% mais pessoas por m^2 sob área útil, e 93% mais, se considerar a área de cobertura das torres de telefonia.

Praias (do Leme ao Pontal)	Visitantes (Média aos domingos de verão)	Pessoas/ m^2 (área útil)	Pessoas/ m^2 (área das torres)
Praia do Leme	34.968	0,158	21,86E-03
Praia de Copacabana	84.746	0,326	58,85E-03
Praia do Arpoador	24.124	0,313	43,86E-03
Praia de Ipanema	28.404	0,154	47,34E-03
Praia do Leblon	25.943	0,187	36,03E-03
Praia de São Conrado	9.731	0,065	7,72E-03
Praia do Pepê (Barra)	34.673	0,178	16,75E-03
Praia do Meio da Barra	30.665	0,186	12,62E-03
Praia da Reserva	86.284	0,261	3,99E-03
Praia do Recreio	33.531	0,289	11,03E-03
Praia do Pontal	14.881	0,187	1,79E-03

Tabela 6.5: Visitantes às praias do Rio de Janeiro no verão 2013/2014, segundo estimativas da metodologia de CDR.

A Figura 6.10 dá enfoque nas regiões da RMRJ que mais produziram viagens às praias. Trata-se da volumetria média da população visitante sobre uma distribuição geográfica que aponta aos locais de origem (residência presumida). O diâmetro dos gráficos de setores está ligado a quantidade de pessoas que não estava no local da própria residência e estava em alguma das praias, exceto para aqueles em que a praia fica na própria região de residência. É sabido que as antenas cobrem regiões difusas, de forma que possivelmente um indivíduo residente na Av. Atlântica (em Copacabana), será contado como visitante, tenha ele ido à praia ou não. Os setores dos gráficos evidenciam as praias escolhidas, de acordo com as cores. O posicionamento dos gráficos está ligado às coordenadas geográficas da região de residência.

O comportamento típico do Reveillon parece repetir-se nas praias, os indivíduos tendem a deslocar-se aos destinos mais próximos. Os residentes da Barra da Tijuca preferem majoritariamente as praias da Reserva, Pepê, Recreio e meio

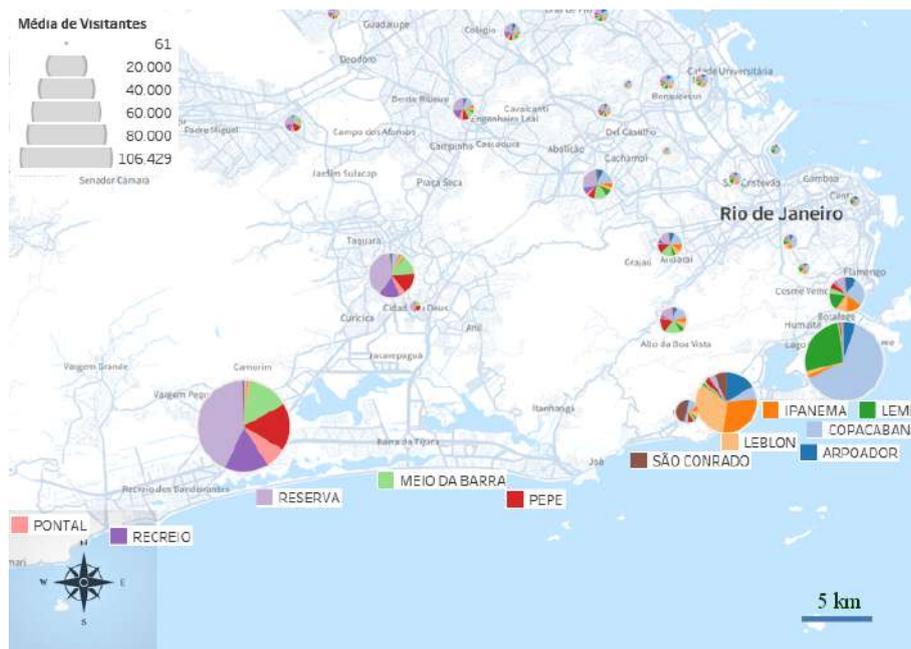


Figura 6.10: Média de visitas às praias da cidade do Rio de Janeiro durante todos os domingos do verão 2013/2014.

da Barra. Já os residentes de Copacabana preferem a praia de Copacabana e uma parcela significativa escolhe a praia do Leme. O mesmo se passa com os residentes do sub-distrito da Lagoa, que preferem em sua maioria Ipanema, Leblon, Arpoador e São Conrado. Em proporções menores, os moradores de Jacarepaguá seguem padrões semelhantes aos moradores da Barra da Tijuca em relação a escolha das praias, ou seja, as praias da Barra. Jacarepaguá está situado a norte da Barra da Tijuca e não possui uma costa com praia.

A Figura 6.11 mostra a probabilidade dos indivíduos escolherem uma determinada praia em função da distância a percorrer. Conforme visto na metodologia, a probabilidade é obtida pela razão entre os indivíduos que realizaram alguma chamada na praia X e todos os usuários que realizaram alguma chamada no intervalo dos domingos de verão, independente da localidade. Cada ponto é um local de origem e cada cor é uma praia, de forma que existem tantos pontos quanto locais de origem x praias. As cores no topo da figura evidenciam as praias que mais atraem a população. Observa-se um padrão difuso que apresenta as praias da Reserva e de Copacabana como as de maior probabilidade. Já na extremidade inferior do gráfico a praia de São Conrado segue um padrão de

baixa probabilidade de visita (pontos vermelhos), ficando difuso a partir de 50 km também.

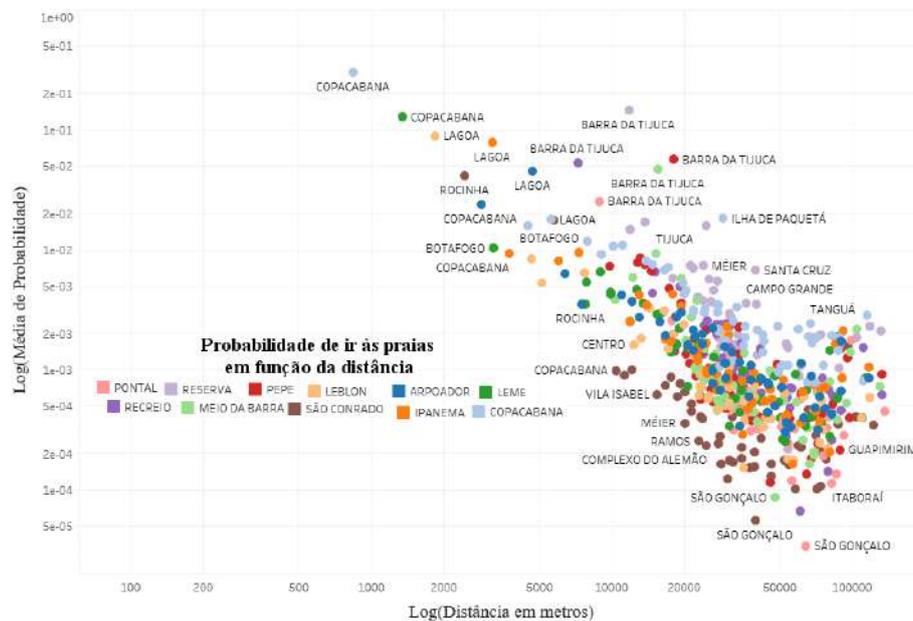


Figura 6.11: Média da probabilidade em função da distância percorrida das pessoas irem às praias da cidade do Rio de Janeiro durante todos os domingos do verão 2013/2014.

Foram mostrados os resultados obtidos para análise da mobilidade da RMRJ por conta do Reveillon, das praias do Rio de Janeiro e do Carnaval. De forma geral, foi visto que a distância regula a disposição de as pessoas irem ou não em um evento, pois as pessoas tendem a ficar mais próximas de casa. Um evento precisa ter um apelo muito forte para que as pessoas percorram longas distâncias, como foi o efeito visto no Reveillon.

Capítulo 7

Conclusão e trabalhos futuros

*“Se depois desta volta para dentro,
deste ensimesmar-se, brotarem
versos, não mais pensará em
perguntas seja a quem for se são
bons. Nem tão pouco tentará
interessar as revistas por esses
trabalhos, pois há de ver neles sua
querida propriedade natural, um
pedaço e uma voz de sua vida.”*

— Rainer Maria Rilke, Cartas a um
jovem poeta.

A característica multidisciplinar deste trabalho fez percorrer por referências de diversas áreas: área de transportes, de ciências sociais, de turismo e da física. As pesquisas foram consistentes em mostrar que o direcionamento do uso de dados de CDR para mobilidade pode ser útil à sociedade, desde que as novas metodologias sejam pautados nos métodos tradicionais da área de transportes.

O estabelecimento de um particionamento geográfico em unidades “conhecidas”, deu suporte a elaboração de fatores de expansão populacional coerentes com o sócio-econômico da população através do órgão oficial do país, o IBGE, e com a rede de torres de telefonia móvel. A título de “evolução” desse particiona-

mento, seria uma boa indicação que algumas cidades que apresentaram grande volumetria de viagens fossem sub-particionadas, como foi o caso de São Gonçalo e de Duque de Caxias.

Foram abordados dois fatores de expansão populacional distintos, de onde se observou que, se comparados às linhas de base do PDTU, um deles superestima a quantidade de viagens (k-adaptativo) e o outro subestima (k-fixo). A recomendação é que sempre se compare os dois fatores aos dados mais atualizados disponibilizados pela área de transportes, sempre indagando a validade de cada um. De forma geral, ao se considerar um projeto de transporte que contemple o “pior cenário” possível, o uso do k-adaptativo seria a indicação mais segura.

Através dos dados de CDR do ano de 2014, de uma operadora de telefonia da RMRJ, foram identificados diversos padrões quotidianos de mobilidade: *a)* casa - trabalho, mostrando comportamento regular em dias úteis; *b)* casa - lazer, mostrando comportamento singular em feriados prolongados; *c)* mudança em padrão, sob possível influência da inauguração do Arco Metropolitano; *d)* possível mudança após implementação do BRT Transcarioca. Assim como, a compreensão de padrões de mobilidade direcionados à eventos. Foram abordados os principais eventos que anualmente ocorrem na RMRJ, tais como Reveillon e Carnaval, além de um mapeamento das preferências dos residentes da RMRJ pelas diversas praias disponíveis na região de estudo.

Os resultados apresentados comprovam a aplicabilidade da utilização de dados de CDR para o planejamento e monitoramento de transportes em grandes cidades, além de permitir outros estudos que não seriam viáveis com as metodologias tradicionais de coleta de dados. O monitoramento da mobilidade durante todo o ano permite o acompanhamento dos resultados da implementação de políticas públicas, assim como o planejamento eficiente de novas intervenções.

A principal limitação desta tese é a dependência do usuário realizar chamadas telefônicas, pois caso ele não as tenha feito, não há como saber a posição em que estava. Uma outra limitação foi encontrada através da tentativa de inferir o local

de trabalho das pessoas, o padrão se mostrou muito difuso e sem sustentação para validação.

O uso de novas bases de dados, tais como: *a)* dados de GPS de celular, *b)* dados da RAIS (Relação Anual de Informações Sociais), *c)* dados de catracas de ônibus (RioCard no caso da RMRJ), seriam uma excelente indicação para que a metodologia fosse mais precisa e refinada.

A título de direcionamento para trabalhos futuros, logo se percebe que a aplicabilidade da metodologia de estimativa de mobilidade vai além da mobilidade em si. Apenas para citar algumas possibilidades, o tema pode seguir para a área da saúde, colaborando em pesquisas sobre o espalhamento de doenças, como é o caso da dengue, por exemplo. De forma aderente e específica à RMRJ, também pode-se estudar a mobilidade entre regiões específicas, conhecidas por altos índices de criminalidade. Um estudo sob esse viés poderia trazer resultados interessantes ao Instituto de Segurança Pública, por exemplo.

O tema específico do fator de expansão adaptativo (Equação 4.5 da Seção 4.3) pode ser revisitado considerando que certa parcela da população é imóvel. Já que, a equação 4.5 considera que 100% da população realiza viagens, o que pode não refletir a realidade. Pode-se até mesmo estimar a população imóvel a partir de comparações com o PDTU.

Referências Bibliográficas

- [1] UNITED NATIONS DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS/POPULATION DIVISION. “World Urbanization Prospects: The 2007 Revision”. 2007.
- [2] MCNALLY, M. G. “The Four-Step Model”. In: Button, K. J., Hensher, D. A. (Eds.), *Handbook of Transport Modelling (Handbooks in Transport, Volume 1)*, 2nd ed., Emerald Group Publishing Limited, cap. 3, pp. 35 – 53, set 2007. ISBN: 978-0-08-045376-7. doi: 10.1108/9780857245670-002.
- [3] COMPANHIA ESTADUAL DE ENGENHARIA DE TRANSPORTE E LOGÍSTICA. *PDTU 2003. Plano Diretor de Transporte Urbano da Região Metropolitana do Rio de Janeiro. Governo do Rio de Janeiro: Secretaria de Transportes. Relatório técnico*, Governo do Rio de Janeiro: Secretaria de Transportes, RJ, 2003.
- [4] COMPANHIA ESTADUAL DE ENGENHARIA DE TRANSPORTE E LOGÍSTICA. *PDTU 2013. Plano Diretor de Transporte Urbano da Região Metropolitana do Rio de Janeiro. Relatório Técnico 4, Planejamento e execução das pesquisas Parte III, Governo do Rio de Janeiro: Secretaria de Transportes. Relatório Técnico 4*, Governo do Rio de Janeiro: Secretaria de Transportes, RJ, 2015.
- [5] JÄRV, O., AHAS, R., WITLOX, F. “Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records”, *Transportation Research Part C: Emerging Technologies*, v. 38, pp. 122–135, 2014. ISSN: 0968090X. doi: 10.1016/j.trc.2013.11.003. Disponível em: <<http://dx.doi.org/10.1016/j.trc.2013.11.003>>.
- [6] CACERES, N., WIDEBERG, J. P., BENITEZ, F. G. “Deriving origin-destination data from a mobile phone network”, *IET INTELLIGENT TRANSPORT SYSTEMS*, v. 1, n. 1, pp. 15–26, mar 2007. ISSN: 1751-956X. doi: 10.1049/iet-its:20060020.

- [7] TOOLE, J. L., ÇOLAK, S., STURT, B., et al. "The path most traveled: Travel demand estimation using big data resources", *Transportation Research Part C: Emerging Technologies*, v. 58, pp. 162–177, 2015. Disponível em: <<http://arxiv.org/abs/1403.0636>>.
- [8] GETZ, D., PAGE, S. J. "Progress and prospects for event tourism research", *Tourism Management*, v. 52, pp. 593–631, 2014. ISSN: 02615177. doi: 10.1016/j.tourman.2015.03.007. Disponível em: <<http://dx.doi.org/10.1016/j.tourman.2015.03.007>>.
- [9] SIMINI, F., GONZÁLEZ, M. C., MARITAN, A., et al. "A universal model for mobility and migration patterns." *Nature*, v. 484, n. 7392, pp. 96–100, feb 2012. ISSN: 1476-4687. doi: 10.1038/nature10856. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22367540>>.
- [10] DAVIS, K. "The Urbanization of the Human Population". 1965. ISSN: 0036-8733.
- [11] SCHRANK., D., EISELE., B., LOMAX., T., et al. *2015 Urban Mobility Scorecard*. Relatório Técnico August, 2015. Disponível em: <<https://d2dt15nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-scorecard-2015.pdf>>.
- [12] "ITU: Committed to connecting the world". 2015. Disponível em: <<http://www.itu.int/en/Pages/default.aspx>>.
- [13] JACOBS, J. *The Death and Life of Great American Cities*. Reissue ed. New York, Vintage, 1992. ISBN: 9780679741954.
- [14] CAREY, H. C. *Principles of Social Science*. Philadelphia, Pennsylvania, J. Lippincott., 1858.
- [15] SEN, A., SMITH, T. *Gravity Models of Spatial Interaction Behavior*. Berlin Heidelberg, Springer-Verlag, 1995. doi: 10.1007/978-3-642-79880-1.
- [16] IVEY, S. S., BADOE, D. A. "Review of Policies on Access to Transportation Planning Data and Models: Implications for Transportation Planning Agencies", *Journal of Urban Planning and Development*, v. 137, n. 4, pp. 438–447, 2011. ISSN: 0733-9488. doi: 10.1061/(ASCE)UP.1943-5444.0000089.
- [17] XAVIER, F. H. Z. *CARACTERIZAÇÃO DA CARGA DE TRABALHO DE UMA REDE DE TELEFONIA MÓVEL EM GRANDES EVENTOS*. Dis-

sertação de mestrado, Pontifícia Universidade Católica, Minas Gerais, 2014.

- [18] JOSÉ, F. F. D. *Análise de grandes eventos na cidade do Rio de Janeiro usando dados de telefonia celular*. Dissertação de mestrado, eMAP/FGV, 2015.
- [19] BATES, J. “History of Demand Modelling”. In: Button, K. J., Hensher, D. A. (Eds.), *Handbook of Transport Modelling (Handbooks in Transport, Volume 1)*, 2nd ed., Emerald Group Publishing Limited, cap. 2, pp. 11–34, set 2007. ISBN: 978-0-08-045376-7. doi: 10.1108/9780857245670-002.
- [20] KIVELÄ, M., ARENAS, A., BARTHELEMY, M., et al. “Multilayer networks”, *Journal of Complex Networks*, v. 2, n. 3, pp. 203–271, 2014. ISSN: 20511329. doi: 10.1093/comnet/cnu016.
- [21] DU, W.-B., ZHOU, X.-L., JUSUP, M., et al. “Physics of transportation : Towards optimal capacity using the multilayer network framework”, *Nature Publishing Group*, v. 6, n. October 2015, pp. 1–8, 2016. ISSN: 2045-2322. doi: 10.1038/srep19059. Disponível em: <<http://dx.doi.org/10.1038/srep19059>>.
- [22] GALLOTTI, R., BARTHELEMY, M. “The multilayer temporal network of public transport in Great Britain”, *Scientific Data*, v. 2, 2015. ISSN: 20524463. doi: 10.1038/sdata.2014.56.
- [23] PAIVA, C. “Modelos Tradicionais Transporte e Tráfego”. Disponível em: <https://www.sinaldetransito.com.br/artigos/modelos_tradicionais.pdf>.
- [24] SANANDAJI, B., VARAIYA, P. “Compressive origin-destination estimation”, *Transportation Letters*, 2016. ISSN: 19427875. doi: 10.1179/1942787515Y.0000000018.
- [25] EL MAHRSI, M. K., COME, E., OUKHELLOU, L., et al. “Clustering Smart Card Data for Urban Mobility Analysis”, *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, v. 18, n. 3, pp. 712–728, mar 2017. ISSN: 1524-9050. doi: 10.1109/TITS.2016.2600515.
- [26] BARCELÓ, J., MONTERO, L. “A robust framework for the estimation of dynamic OD trip matrices for reliable traffic management”. In: *Transportation Research Procedia*, 2015. ISBN: 9781467365956. doi: 10.1016/j.trpro.2015.09.063.

- [27] MARTIN, W. A., MCGUCKIN, N. A. *Travel Estimation Techniques for Urban Planning*. NCHRP 365, National Cooperative Highway Research Program, Washington, DC, 1998.
- [28] ORTUZAR S., J. D. D., WILLUMSEN, L. G. *Modelling transport*. UK, John Wiley & Sons, Inc., 2011. ISBN: 0470760397.
- [29] ECONOMIC, U. N., AFFAIRS, S. "The World's Cities in 2016 - Data Booklet". 2016. Disponível em: <<http://www.un.org/en/development/desa/population/publications/>>.
- [30] BUREAU OF URBAN DEVELOPMENT TOKYO METROPOLITAN GOVERNMENT. "Planning Tokyo's Urban Development". 2016. Disponível em: <<http://www.toshiseibi.metro.tokyo.jp/eng/>>.
- [31] MINISTRY OF URBAN DEVELOPMENT, I. "Functional Plan on Transport for National Capital Region-2032". 2016. Disponível em: <<http://ncrpb.nic.in>>.
- [32] SÃO PAULO TRANSPORTES S.A., COMPANHIA DE ENGENHARIA DE TRÁFEGO. "Plano de Mobilidade de São Paulo". 2015. Disponível em: <<http://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/planmob/>>.
- [33] HOLLINGSHEAD, A. B., ZIPF, G. K. "Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology." 1949. ISSN: 00031224.
- [34] BLONDEL, V. D., DECUYPER, A., KRINGS, G. "A survey of results on mobile phone datasets analysis", *EPJ Data Science*, v. 4, n. 10, dec 2015. ISSN: 2193-1127. doi: 10.1140/epjds/s13688-015-0046-0. Disponível em: <<http://www.epjdatascience.com/content/4/1/10>>.
- [35] CHEN, C., MA, J., SUSILO, Y., et al. "The promises of big data and small data for travel behavior (aka human mobility) analysis", *Transportation Research Part C: Emerging Technologies*, v. 68, pp. 285–299, jul 2016. ISSN: 0968-090X. doi: 10.1016/J.TRC.2016.04.005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0968090X16300092>>.
- [36] HOTEIT, S., SECCI, S., SOBOLEVSKY, S., et al. "Estimating human trajectories and hotspots through mobile phone data", *Computer Networks*, v. 64, pp. 296–307, 2014. ISSN: 13891286. doi: 10.1016/j.comnet.2014.02.011.

- [37] KWOK, R. “Personal technology: Phoning in data.” *Nature*, v. 458, n. 7241, pp. 959–961, 2009. ISSN: 0028-0836. doi: 10.1038/458959a.
- [38] TRESTIAN, R., SHAH, P., NGUYEN, H., et al. “Towards connecting people, locations and real-world events in a cellular network”, *Telematics and Informatics*, v. 34, n. 1, pp. 244–271, 2017. ISSN: 07365853. doi: 10.1016/j.tele.2016.05.006.
- [39] ZHONG, E., TAN, B., MO, K., et al. “User demographics prediction based on mobile data”, *Pervasive and Mobile Computing*, v. 9, n. 6, pp. 823–837, dec 2013. ISSN: 15741192. doi: 10.1016/j.pmcj.2013.07.009.
- [40] DONG, H., WU, M., DING, X., et al. “Traffic zone division based on big data from mobile phone base stations”, *Transportation Research Part C: Emerging Technologies*, v. 58, pp. 278–291, 2015. ISSN: 0968090X. doi: 10.1016/j.trc.2015.06.007.
- [41] SMOREDA, Z., OLTEANU-RAIMOND, A.-M., COURONNÉ, T. “Spatiotemporal Data from Mobile Phones for Personal Mobility Assessment”. In: *Transport Survey Methods: Best Practice for Decision Making*, Emerald Group Publishing Limited, pp. 745–767, jan 2013. doi: doi: 10.1108/9781781902882-041. Disponível em: <<http://dx.doi.org/10.1108/9781781902882-041>>.
- [42] YANG, Y., JIANG, S., VENEZIANO, D., et al. “TimeGeo: modeling urban mobility without travel surveys”, *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, n. 51, pp. 20167–20172, 2007. ISSN: 0027-8424. doi: 10.1073/pnas.0709640104.
- [43] LIU, F., JANSSENS, D., WETS, G., et al. “Annotating mobile phone location data with activity purposes using machine learning algorithms”, *Expert Systems with Applications*, v. 40, n. 8, pp. 3299–3311, 2013. ISSN: 09574174. doi: 10.1016/j.eswa.2012.12.100.
- [44] ALEXANDER, L., JIANG, S., MURGA, M., et al. “Origin-destination trips by purpose and time of day inferred from mobile phone data”, *Transportation Research Part C: Emerging Technologies*, v. 58, pp. 240–250, 2015. ISSN: 0968090X. doi: 10.1016/j.trc.2015.02.018. Disponível em: <<http://dx.doi.org/10.1016/j.trc.2015.02.018>>.
- [45] OKABE, A., BOOTS, B., SUGIHARA, K., et al. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. 2nd ed. New York, Wiley, 2000. ISBN: 978-0-471-98635-5.

- [46] CANDIA, J., GONZÁLEZ, M. C., WANG, P., et al. “Uncovering individual and collective human dynamics from mobile phone records”, *Journal of Physics A: Mathematical and Theoretical*, v. 41, n. 22, pp. 224015, 2008. ISSN: 1751-8113. doi: 10.1088/1751-8113/41/22/224015.
- [47] DEVILLE, P., LINARD, C., MARTIN, S., et al. “Dynamic population mapping using mobile phone data”, *Proceedings of the National Academy of Sciences*, 2014. ISSN: 0027-8424. doi: 10.1073/pnas.1408439111.
- [48] DOBSON, J. E., BRIGHT, E. A., COLEMAN, P. R., et al. “LandScan: A global population database for estimating populations at risk”, *Photogrammetric Engineering and Remote Sensing*, 2000. ISSN: 0099-1112. doi: 10.1016/j.scitotenv.2008.02.010.
- [49] WANG, L., HU, K., KU, T., et al. “Urban Mobility Dynamics Based on Flexible Discrete Region Partition”, *INTERNATIONAL JOURNAL OF DISTRIBUTED SENSOR NETWORKS*, 2014. ISSN: 1550-1329. doi: 10.1155/2014/782649.
- [50] SCHNEIDER, C. M., BELIK, V., COURONNÉ, T., et al. “Unravelling daily human mobility motifs.” *Journal of The Royal Society Interface*, v. 10, n. 20130246, 2013. ISSN: 1742-5662. doi: 10.1098/rsif.2013.0246.
- [51] JIANG, S., FERREIRA JR, J., GONZÁLEZ, M. C. “Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore”. In: *The 4th International Workshop on Urban Computing*, 2015.
- [52] MOREIRA-MATIAS, L., GAMA, J., FERREIRA, M., et al. “Time-evolving O-D matrix estimation using high-speed GPS data streams”, *Expert Systems with Applications*, 2015. ISSN: 09574174. doi: 10.1016/j.eswa.2015.08.048.
- [53] IQBAL, M. S., CHOUDHURY, C. F., WANG, P., et al. “Development of origin-destination matrices using mobile phone call data”, *Transportation Research Part C: Emerging Technologies*, v. 40, pp. 63–74, 2014. ISSN: 0968090X. doi: 10.1016/j.trc.2014.01.002. Disponível em: <<http://dx.doi.org/10.1016/j.trc.2014.01.002>>.
- [54] DEMISSIE, M. G., PHITHAKKITNUKON, S., SUKHVIBUL, T., et al. “Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal”, *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION*

SYSTEMS, v. 17, n. 9, pp. 2466–2478, 2016. ISSN: 1524-9050. doi: 10.1109/TITS.2016.2521830.

- [55] GETZ, D. “Event tourism: Definition, evolution, and research”, *TOURISM MANAGEMENT*, v. 29, n. 3, pp. 403–428, jun 2008. ISSN: 0261-5177. doi: 10.1016/j.tourman.2007.07.017.
- [56] PARKES, S. D., JOPSON, A., MARSDEN, G. “Understanding travel behaviour change during mega-events: Lessons from the London 2012 Games”, *Transportation Research Part A: Policy and Practice*, pp. 104–119, 2016. ISSN: 09658564. doi: 10.1016/j.tra.2016.07.006.
- [57] MA, S.-C., EGAN, D., ROTHERHAM, I., et al. “A framework for monitoring during the planning stage for a sports mega-event”, *Journal of Sustainable Tourism*, v. 19, n. 1, pp. 79–96, 2011. ISSN: 0966-9582. doi: 10.1080/09669582.2010.502576.
- [58] QUINTELLA, C. O. D. V. *Segurança em Jogo – Diagnóstico em matéria de Segurança Pública face à vocação do Rio de Janeiro em sediar Eventos de Grande Porte*. Dissertação de mestrado, EBAPE/FGV, 2011.
- [59] XAVIER, F. H. Z., SILVEIRA, L. M., DE ALMEIDA, J. M., et al. “Analyzing the workload dynamics of a mobile phone network in large scale events”. In: *Proceedings of the first workshop on Urban networking - UrbaNe '12*, pp. 37–42, Nice, France, 2012. ACM. ISBN: 9781450317818. doi: 10.1145/2413236.2413245. Disponível em: <<http://dl.acm.org/citation.cfm?id=2413236.2413245>>.
- [60] XAVIER, F. H. Z., SILVEIRA, L., ALMEIDA, J., et al. “Understanding human mobility due to large-scale events”. In: *Third international conference on the analysis of mobile phone datasets (NetMob)*, 2013.
- [61] CALABRESE, F., COLONNA, M., LOVISOLO, P., et al. “Real-Time Urban Monitoring Using Cellular Phones: a Case-Study in Rome”, *IEEE Transactions on Intelligent Transportation Systems*, v. 12, n. 1, pp. 1–11, 2010. ISSN: 1524-9050. doi: 10.1109/tits.2010.2074196.
- [62] BETTENCOURT, L. M. A., LOBO, J., HELBING, D., et al. “Growth, innovation, scaling, and the pace of life in cities.” *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, n. 17, pp. 7301–7306, 2007. ISSN: 0027-8424. doi: 10.1073/pnas.0610172104.

- [63] WESOŁOWSKI, A., O’MEARA, W. P., EAGLE, N., et al. “Evaluating Spatial Interaction Models for Regional Mobility in Sub-Saharan Africa”, *PLoS Computational Biology*, v. 11, n. 7, pp. 1–16, 2015. ISSN: 15537358. doi: 10.1371/journal.pcbi.1004267.
- [64] MASUCCI, A. P., SERRAS, J., JOHANSSON, A., et al. “Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows”, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, v. 88, n. 2, 2013. ISSN: 15393755. doi: 10.1103/PhysRevE.88.022812.
- [65] BALCAN, D., COLIZZA, V., GONÇALVES, B., et al. “Multiscale mobility networks and the spatial spreading of infectious diseases”, *Proceedings of the National Academy of Sciences*, v. 106, n. 51, pp. 21484–21489, 2009. ISSN: 0027-8424. doi: 10.1073/pnas.0906910106. Disponível em: <<http://www.pnas.org/content/106/51/21484>>.
- [66] ONNELA, J.-P., SARAMÄKI, J., HYVÖNEN, J., et al. “Structure and tie strengths in mobile communication networks”, *Proceedings of the National Academy of Sciences*, v. 104, n. 18, pp. 7332 LP – 7336, may 2007. Disponível em: <<http://www.pnas.org/content/104/18/7332.abstract>>.
- [67] NEWMAN, M. E. J. *Networks: An introduction*. 2014. ISBN: 9783319035178. doi: 10.1007/978-3-319-03518-5-8.
- [68] BANKIER, M. D. “Generalized Least Squares Estimation Under Poststratification”. In: *Proceeding of the Section on Research Methods*, pp. 730–755. American Statistical Association, 1990.
- [69] LOUAIL, T., LENORMAND, M., CANTU ROS, O. G., et al. “From mobile phone data to the spatial structure of cities”, *Scientific Reports*, v. 4, 2014. ISSN: 20452322. doi: 10.1038/srep05276.
- [70] BARLACCHI, G., DE NADAI, M., LARCHER, R., et al. “A multi-source dataset of urban life in the city of Milan and the Province of Trentino”, *Scientific Data*, v. 2, pp. 150055, oct 2015. ISSN: 2052-4463. doi: 10.1038/sdata.2015.55.
- [71] DOYLE, J. *Estimating Movement from Mobile Telephony Data*. Tese de Doutorado, National University of Ireland Maynooth, 2014. Disponível em: <<http://eprints.maynoothuniversity.ie/5400/>>.

- [72] JACOBS, H. A. “To count a crowd”, *Columbia Journalism Review*, v. 6, pp. 36–40, 1967.
- [73] CHOI-FITZPATRICK, A., JUSKAUSKAS, T. “Up in the Air: Applying the Jacobs Crowd Formula to Drone Imagery”, *Procedia Engineering*, v. 107, pp. 273–281, 2015. ISSN: 18777058. doi: 10.1016/j.proeng.2015.06.082. Disponível em: <<http://dx.doi.org/10.1016/j.proeng.2015.06.082>>.
- [74] NAINI, F. M., DOUSSE, O., THIRAN, P., et al. “Opportunistic Sampling for Joint Population Size and Density Estimation”, *IEEE Transactions on Mobile Computing*, v. 14, n. 12, pp. 2530–2543, 2015. ISSN: 15361233. doi: 10.1109/TMC.2015.2393302.
- [75] BONELLI, M. C. *Sustentabilidade em Obras Públicas: O Caso do Parque Madureira*. Dissertação de mestrado, PUC-Rio Pontifícia Universidade Católica do Rio de Janeiro, 2013.
- [76] LAMB, A., FULLER, M., VARADARAJAN, R., et al. “The Vertica Analytic Database: C-Store 7 Years Later”, *Proceedings of the VLDB Endowment*, v. 5, n. 12, pp. 1790–1801, 2012. ISSN: 2150-8097. Disponível em: <<http://arxiv.org/abs/1208.4173>>.
- [77] STONEBRAKER, M., ABADI, D. J., BATKIN, A., et al. “C-store: a column-oriented DBMS”. In: *VLDB Conference*, pp. 553–564, Trondheim, 2005. ISBN: 1-59593-154-6. doi: 10.1007/BF02443652.
- [78] TUKEY, J. W., TUKEY, P. “Computer graphics and exploratory data analysis: An introduction”, *Sixth Annual Conference and Exposition: Computer Graphics*, pp. 773–785, 1985.
- [79] UNWIN, A., THEUS, M., HOFMANN, H. *Graphics of Large Datasets: Visualizing a Million (Statistics and Computing)*. Secaucus, NJ, USA, Springer-Verlag New York, Inc., 2006. ISBN: 0387329064.
- [80] YAU, N. *Visualize this: the FlowingData guide to design, visualization, and statistics*. Indianapolis, Ind, Wiley Pub, 2011. ISBN: 9780470944882.
- [81] “Nokia Mobile data challenge”. 2012. Disponível em: <<http://research.nokia.com/page/12000>>.
- [82] SLINGSBY, A., BEECHAM, R., WOOD, J. “Visual analysis of social networks in space and time using smartphone logs”, *Pervasive and Mobile Com-*

puting, v. 9, n. 6, pp. 848–864, dec 2013. ISSN: 15741192. doi: 10.1016/j.pmcj.2013.07.002.

[83] SLINGSBY, A. “Supporting the visual analysis of the behaviour of gulls”, 2011. Disponível em: <<http://openaccess.city.ac.uk/2528/>>.

[84] SLINGSBY, A., DYKES, J. “Experiences in involving analysts in visualisation design”, *Proceedings of the 2012 BELIV Workshop on Beyond Time and Errors - Novel Evaluation Methods for Visualization - BELIV '12*, pp. 1–4, 2012. doi: 10.1145/2442576.2442577.

Apêndice A

Apêndice

A.1 Unidades geográficas de estudo

Tabela A.1: Unidades do particionamento misto adotadas para a área de estudo, relacionadas às macrozonas de tráfego da RMRJ.

Unidade Geográfica desta pesquisa	Macrozona de tráfego	Dentro da RMRJ?
ANCHIETA	AP3	<input checked="" type="checkbox"/>
BANGU	AP5	<input checked="" type="checkbox"/>
BARRA DA TIJUCA	AP4	<input checked="" type="checkbox"/>
BELFORD ROXO	BAIXADA 1B	<input checked="" type="checkbox"/>
BOTAFOGO	AP2	<input checked="" type="checkbox"/>
CACHOEIRAS DE MACACU	EXTRA URBS	<input type="checkbox"/>
CAMPO GRANDE	AP5	<input checked="" type="checkbox"/>
CENTRO	AP1	<input checked="" type="checkbox"/>
CIDADE DE DEUS	AP4	<input checked="" type="checkbox"/>
COMPLEXO DO ALEMÃO	AP3	<input checked="" type="checkbox"/>
COPACABANA	AP2	<input checked="" type="checkbox"/>
DUQUE DE CAXIAS	BAIXADA 2	<input checked="" type="checkbox"/>
GUAPIMIRIM	BAIXADA 2	<input checked="" type="checkbox"/>
GUARATIBA	AP5	<input checked="" type="checkbox"/>
ILHA DE PAQUETÁ	AP1	<input checked="" type="checkbox"/>
ILHA DO GOVERNADOR	AP3	<input checked="" type="checkbox"/>
INHAÚMA	AP3	<input checked="" type="checkbox"/>
IRAJÁ	AP3	<input checked="" type="checkbox"/>
ITABORAÍ	LESTE METROPOLITANO 2	<input checked="" type="checkbox"/>
ITAGUAÍ	SUL METROPOLITANO	<input checked="" type="checkbox"/>
JACAREPAGUÁ	AP4	<input checked="" type="checkbox"/>
JACAREZINHO	AP3	<input checked="" type="checkbox"/>
JAPERI	BAIXADA 1A	<input checked="" type="checkbox"/>
LAGOA	AP2	<input checked="" type="checkbox"/>
MADUREIRA	AP3	<input checked="" type="checkbox"/>
MAGÉ	BAIXADA 2	<input checked="" type="checkbox"/>

Tabela A.1: Unidades do particionamento misto adotadas para a área de estudo, relacionadas às macrozonas de tráfego da RMRJ.

Unidade Geográfica desta pesquisa	Macrozona de tráfego	Dentro da RMRJ?
MANGARATIBA	SUL METROPOLITANO	<input checked="" type="checkbox"/>
MARICÁ	LESTE METROPOLITANO 1	<input checked="" type="checkbox"/>
MARÉ	AP3	<input checked="" type="checkbox"/>
MESQUITA	BAIXADA 1B	<input checked="" type="checkbox"/>
MÉIER	AP3	<input checked="" type="checkbox"/>
NILÓPOLIS	BAIXADA 1B	<input checked="" type="checkbox"/>
NITERÓI	LESTE METROPOLITANO 1	<input checked="" type="checkbox"/>
NOVA IGUAÇU	BAIXADA 1A	<input checked="" type="checkbox"/>
PARACAMBI	SUL METROPOLITANO	<input checked="" type="checkbox"/>
PAVUNA	AP3	<input checked="" type="checkbox"/>
PENHA	AP3	<input checked="" type="checkbox"/>
PORTUÁRIA	AP1	<input checked="" type="checkbox"/>
QUEIMADOS	BAIXADA 1A	<input checked="" type="checkbox"/>
RAMOS	AP3	<input checked="" type="checkbox"/>
REALENGO	AP5	<input checked="" type="checkbox"/>
RIO BONITO	EXTRA URBS	<input type="checkbox"/>
RIO COMPRIDO	AP1	<input checked="" type="checkbox"/>
ROCINHA	AP2	<input checked="" type="checkbox"/>
SANTA CRUZ	AP5	<input checked="" type="checkbox"/>
SANTA TERESA	AP1	<input checked="" type="checkbox"/>
SEROPÉDICA	SUL METROPOLITANO	<input checked="" type="checkbox"/>
SÃO CRISTOVÃO	AP1	<input checked="" type="checkbox"/>
SÃO GONÇALO	LESTE METROPOLITANO 2	<input checked="" type="checkbox"/>
SÃO JOÃO DE MERITI	BAIXADA 1B	<input checked="" type="checkbox"/>
TANGUÁ	LESTE METROPOLITANO 2	<input checked="" type="checkbox"/>
TERESÓPOLIS	EXTRA URBS	<input type="checkbox"/>
TIJUCA	AP2	<input checked="" type="checkbox"/>
VIGÁRIO GERAL	AP3	<input checked="" type="checkbox"/>

Tabela A.1: Unidades do particionamento misto adotadas para a área de estudo, relacionadas às macrozonas de tráfego da RMRJ.

Unidade Geográfica desta pesquisa	Macrozona de tráfego	Dentro da RMRJ?
VILA ISABEL	AP2	☒
São 55 unidades	12 macrozonas	52 unidades na RMRJ

A.2 Instruções SQL ANSI

A metodologia foi desenvolvida em códigos SQL. Esta seção foi criada para evitar o texto com códigos imensos durante as explicações. Observe que no curso do texto, no capítulo de metodologia, alguns nomes podem ter sido mudados para auxiliar a compreensão. Abaixo, os nomes podem estar “menos amigáveis”, sob o formato mais próximo do original.

A.2.1 Extração de locais de permanência dos usuários

SQL 4.3.b

Entrada: Usuários significativos, posições das torres de telefonia e CDRs.

Parâmetro: Domingos, feriados e intervalo temporal do que se considera que a maioria das pessoas esteja em casa.

Saida: Lista dos locais (PARTICAO_GEOGRAFIA) em que um usuário mais foi “detectado”, segundo os parâmetro estabelecidos.

Extração de local de permanência

```

1 CREATE TABLE ultima_posicao_dias
2 AS
3 SELECT a.ID,
4 dt_ini_atividade_rede AS dia,
5 MAX (hr_ini_atividade_rede) AS MAX_HORA,
```

```

6     PARTICAO_GEOGRAFIA
7 FROM CDR a
8     INNER JOIN LAT_LONG_LOCAL b
9 ON a.longitude = b.longitude AND a.latitude = b.latitude
10    INNER JOIN ligadores_significativos c
11 ON c.ID = a.ID
12 HERE -- qualquer hora em feriados
13 dt_ini_atividade_rede IN (SELECT dia
14                            FROM feriado)
15 OR -- qualquer hora em domingos
16 EXTRACT (dow FROM dt_ini_atividade_rede) = 7
17 OR -- apenas entre 19h e 7h em dias de semana
18 ( EXTRACT (HOUR FROM hr_ini_atividade_rede) > 19
19 OR EXTRACT (HOUR FROM hr_ini_atividade_rede) < 6)
20 GROUP BY a.ID,
21 dt_ini_atividade_rede,
22 PARTICAO_GEOGRAFIA;

```

A.2.2 Quantidade de dias no local mais visitado

SQL 4.3.c1

Entrada: Lista de usuários e todos os locais por ele frequentados.

Parâmetro: -

Saida: Quantidade de dias que o usuário foi encontrado no local em que mais frequentou.

```

_____ Total de dias em locais _____
1 CREATE TABLE contar_dias_em_locais
2 AS
3 SELECT ID, SUM (n_dias) AS total_dias_em_locais
4 FROM ( SELECT ID,
5         PARTICAO_GEOGRAFIA,

```

```
6     COUNT (DIA) AS n_dias
7 FROM ultima_posicao_dias
8     GROUP BY ID, PARTICAO_GEOGRAFIA)
9     AS ultima_posicao
10    GROUP BY ID
11    order by ID, total_dias_em_locais;
```

A.2.3 Cálculo de residência presumida

SQL 4.3.c2

Entrada: Total de dias em locais mais visitados pelos usuários.

Parâmetro: Quantidade proporcional de dias no local mais visitado.

Saida: Cálculo do local de residência presumida dos usuários.

Residência presumida dos usuários

```

1 CREATE TABLE residencia_presumida
2 AS
3 WITH residencia_presumida_bruta
4 AS (SELECT ID, PARTICAO_GEOGRAFIA, MAX (n_dias) AS dias_local_mais_visitado
5 FROM ( -- Conta a quantidade de dias por local de permanencia por ID
6 SELECT ID, PARTICAO_GEOGRAFIA, COUNT (DIA) AS n_dias
7 FROM
8 ultima_posicao_dias
9 GROUP BY ID, PARTICAO_GEOGRAFIA)
10 AS ultima_posicao
11 group by ID, PARTICAO_GEOGRAFIA)
12 select ID, PARTICAO_GEOGRAFIA, dias_local_mais_visitado, total_dias_em_locais
13 FROM
14 (
15 SELECT T1.ID,
16 T1.PARTICAO_GEOGRAFIA,
17 T1.dias_local_mais_visitado,
18 T2.total_dias_em_locais
19 FROM residencia_presumida_bruta T1
20 INNER JOIN contar_dias_em_locais T2 ON T2.ID=T1.ID
21 ) AS residencia_presumida_validacao
22 where total_dias_em_locais >= 8
23 and dias_local_mais_visitado > total_dias_em_locais / 2;

```

A.2.4 Fator de expansão populacional fixo

SQL 4.3.1.1.c

Entrada: Local de residência presumida dos usuários e base de população do IBGE.

Parâmetro: -

Saida: Fator de expansão populacional fixo.

Fator de expansão populacional fixo

```

1   create table fator_k_l as
2   select
3   r.PARTICAO_GEOGRAFIA,
4   count(r.ID) as USU,
5   POP,
6   POP/count(r.ID) as k_l
7   from
8   residencia_presumida r
9   inner join base_IBGE i
10  on i.PARTICAO_GEOGRAFIA = r.PARTICAO_GEOGRAFIA
11  group by r.PARTICAO_GEOGRAFIA, POP;

```

A.2.5 Fator de expansão populacional adaptativo

SQL 4.3.1.2.e

Entrada: Viagens “puras”, ou seja, sem fator de expansão algum.

Parâmetro: -

Saida: Viagens realizadas por dia pelos residentes de l .

Viagens realizadas por dia pelos residentes de l

```

1   create table julio.viajantes_dia_l as
2   select cast(dia as date) as dia,
3   local_residência, sum(viagens) as viagens
4   from julio.matriz_od_transiente_viagens_purasn
5   group by dia, local_residência order by dia, local_residência;

```

Entrada: Viagens realizadas por dia pelos residentes de l .

Parâmetro: -

Saida: #Viagens expandidas adaptativamente à OD e à d .

Viagens expandidas adaptativamente

```

1 create table julio.viagens_adaptativas as
2 select
3 dia, origem, destino, sum(viagens_OD_feitas_por_l) as viagens_feitas_por_l,
4 sum(viagens_adaptativas) as viagens_adaptativas
5 from (
6 select
7 a.dia,
8 a.local_residência as l,
9 a.origem, a.destino,
10 a.viagens as viagens_OD_feitas_por_l,
11 b.viagens as total_de_viagens_d_l,
12 c.pop_censo as pop_l,
13 ceil((a.viagens/b.viagens)*(c.pop_censo)) as viagens_adaptativas
14 from
15 julio.matriz_od_transiente_viagens_purasn a inner join
16 julio.viajantes_dia_l b on a.dia=b.dia and
17 b.local_residência=a.local_residência inner join
18 julio.local_pop_usu c on c.nm_grao = a.local_residência
19 order by
20 a.dia, a.local_residência,
21 a.origem,
22 a.destino
23 ) as s0
24 group by dia, origem, destino
25 order by dia, origem, destino;

```

A.2.6 Viagens por dia

SQL 4.4.1

Entrada: Local de residência presumida dos usuários, base de CDR e base de torres de telefonia.

Parâmetro: O dia, o intervalo temporal entre as chamadas e a distância percorrida.

Saida: Quantidade de deslocamentos entre todas as origens e destino por dia.

```

_____ Viagens por dia com k-fix _____
1  insert into julio.deslocamentos_dia_k
2  SELECT '2014-01-01' AS dia,
3  origem,
4  destino,
5  sum(deslocamentos) as deslocamentos
6  FROM (
7  SELECT
8  origem,
9  destino,
10 Count(*) as deslocamentos
11 FROM (SELECT id,
12 datahora,
13 salto,
14 origem,
15 destino,
16 ocoord,
17 dcoord,
18 minsbetween
19 FROM (SELECT id,
20 datahora,
21 salto,
22 origem,
23 destino,
24 ocoord,
25 dcoord,
26 Round(Distance(Cast(Substr(ocoord, 0, Instr(ocoord, ',')))
27 AS
```

```

28         FLOAT), Cast
29     (
30         Substr(ocoord, Instr(ocoord, ',') + 1,
31         Length(ocoord))AS
32         FLOAT),
33     Cast(
34         Substr(dcoord, 0, Instr(dcoord, ','))AS FLOAT
35     ), Cast
36     (
37         Substr(dcoord, Instr(dcoord, ',') + 1,
38         Length(dcoord))AS
39         FLOAT)), 3)
40         AS km_distance,
41     Datediff(mi, hora, nexthour) AS minsbetween
42 FROM (SELECT id,
43         datahora,
44         salto,
45         hora,
46         Lead(hora, 1)
47         OVER (
48             partition BY id
49             ORDER BY datahora) AS nexthour,
50         local          AS Origem,
51         Lead(local, 1)
52         OVER (
53             partition BY id
54             ORDER BY datahora) AS Destino,
55         coord          AS OCoord,
56         Lead(coord, 1)
57         OVER (
58             partition BY id

```

```

59         ORDER BY datahora) AS DCoord
60     FROM (SELECT ddd_orig
61             || id1          AS id,
62             Upper(b.nm_grao) AS local,
63             b.latitude
64             || ','
65             || b.longitude AS coord,
66             dt_ini_atividade_rede
67             || '_'
68             || hr_ini_atividade_rede AS datahora,
69             hr_ini_atividade_rede AS hora,
70             Conditional_change_event(b.latitude
71                                     || ','
72                                     || b.longitude)
73         OVER (
74             partition BY ddd_orig||id1
75             ORDER BY ddd_orig||id1,
76             dt_ini_atividade_rede||'_'||
77             hr_ini_atividade_rede)
78             AS salto
79     FROM bill_quando_onde_j a
80         INNER JOIN julio.lat_long_local b
81             ON a.longitude = b.longitude
82             AND a.latitude = b.latitude
83         LEFT OUTER JOIN julio.residencia_presumida r
84             ON r.cod = a.ddd_orig||a.id1
85         INNER JOIN julio.local_pop_usu c
86             ON c.nm_grao = r.nm_grao
87     WHERE dt_ini_atividade_rede = '2014-01-01'
88     ORDER BY ddd_orig
89             || id1,

```

```

90         salto,
91         dt_ini_atividade_rede
92         || ' '
93         || hr_ini_atividade_rede,
94         b.nm_grao) AS salto)AS grouped)AS
95 grouped2
96 WHERE ocoord != dcoord
97 AND km_distance > 2
98 AND minsbetween > 30
99 AND minsbetween < 240) AS grouped3
100 GROUP BY origem,
101         destino, k) as LAST
102 GROUP BY origem, destino;

```

SQL 4.4.1.d.1

Entrada: Local de residência presumida dos usuários, base de CDR e base de torres de telefonia e fatores de expansão fixo.

Parâmetro: -

Saida: Quantidade de deslocamentos da população, entre todas as origens e destino por dia.

Viagens por dia com k-fixo

```

1  SELECT
2  origem,
3  destino,
4  Count(*) * k_fixo as viagens_k_fixo,
5  dia
6  FROM ...
7  ... SELECT
8  ...
9         c.k           AS k_fixo,
10        dt_ini_atividade_rede AS dia,

```

```

11      ...
12      FROM CDR a
13           INNER JOIN lat_long_local b
14                ON a.longitude = b.longitude
15                AND a.latitude = b.latitude
16           INNER JOIN residencia_presumida r
17                ON r.ID = a.ID
18           INNER JOIN fator_k_l c
19                ON c.PARTICAO_GEOGRAFIA =
20                r.PARTICAO_GEOGRAFIA

```

SQL 4.4.1.d.2

Entrada: Local de residência presumida dos usuários, base de CDR e base de torres de telefonia e fatores de expansão adaptativo.

Parâmetro: -

Saida: Quantidade de deslocamentos da população, entre todas as origens e destino por dia.

————— Viagens por dia com k-adaptativo —————

```

1  SELECT
2  origem,
3  destino,
4  Count(*) * k_adaptativo as viagens_k_adaptativo,
5  dia
6  FROM ...
7  ... SELECT
8  ...
9           c.k           AS k_adaptativo,
10          dt_ini_atividade_rede AS dia,
11  ...
12  FROM CDR a
13          INNER JOIN lat_long_local b

```

```

14         ON a.longitude = b.longitude
15         AND a.latitude = b.latitude
16     INNER JOIN residencia_presumida r
17         ON r.ID = a.ID
18     INNER JOIN fator_k_l_d c
19         ON c.PARTICAO_GEOGRAFIA =
20         r.PARTICAO_GEOGRAFIA
21         and c.DIA = a.dt_ini_atividade_rede

```

SQL 4.5.c.1

Entrada: Local de residência presumida dos usuários, base de CDR e base de torres de telefonia.

Parâmetro: O intervalo espaço-temporal do evento, através das tabelas EVENTO e EVENTO_LATLNG.

Saida: Descoberta de usuários que estavam no evento.

```

_____ Usuários no evento _____
1  create table EVENTO_REVEILLON2014_DENTRO as
2  select ID , a.latlng
3  from CDR a inner join EVENTO_LATLNG b on
4  a.LAT_LNG=b.LAT_LNG inner join EVENTO c on
5  b.ID_EVENTO=c.ID_EVENTO inner join residencia_presumida d on
6  d.ID = a.ID
7  where
8  ((
9  DT_INI_ATIVIDADE_REDE in (
10 ast(date_part('year',c.inicio)||'-'||date_part('month',c.inicio)||'-'||
11 ate_part('day',c.inicio) as date)
12
13 nd HR_INI_ATIVIDADE_REDE > cast(date_part('hour',c.inicio)||':'||
14 ate_part('minute',c.inicio)||':'||date_part('second',c.inicio) as time)
15 )

```

```

16 and
17 (
18     DT_INI_ATIVIDADE_REDE in (
19     ast(date_part('year',c.fim)||'-'||date_part('month',c.fim)||'-'||
20     ate_part('day',c.fim) as date)
21
22     nd HR_INI_ATIVIDADE_REDE < cast(date_part('hour',c.fim)||':'||
23     ate_part('minute',c.fim)||':'||date_part('second',c.fim) as time)
24 ))
25 and b.ID_EVENTO = 'REVEILLON2014'
26 group by ID, a.LAT_LNG;

```

SQL 4.5.c.2

Entrada: Local de residência presumida dos usuários, base de CDR e base de torres de telefonia.

Parâmetro: O intervalo espaço-temporal do evento, através das tabelas EVENTO e EVENTO_LATLNG.

Saida: Descoberta de usuários que estavam no evento.

```

_____ Usuários fora do evento _____
1 create table EVENTO_REVEILLON2014_FORA as
2 select a.ID , a.latlng
3 from CDR a FULL OUTER join EVENTO_LATLNG b on
4 .LAT_LNG=b.LAT_LNG inner join EVENTO c on
5 .ID_EVENTO=c.ID_EVENTO inner join residencia_presumida d on
6 .ID = a.ID
7 where
8 ((
9 T_INI_ATIVIDADE_REDE in (
10 cast(date_part('year',c.inicio)||'-'||date_part('month',c.inicio)||'-'||
11 date_part('day',c.inicio) as date)
12 )

```

```

13 and HR_INI_ATIVIDADE_REDE > cast(date_part('hour',c.inicio)||':'||
14 date_part('minute',c.inicio)||':'||date_part('second',c.inicio) as time)
15 )
16 and
17 (
18 T_INI_ATIVIDADE_REDE in (
19 cast(date_part('year',c.fim)||'-'||date_part('month',c.fim)||'-'||
20 date_part('day',c.fim) as date)
21 )
22 and HR_INI_ATIVIDADE_REDE < cast(date_part('hour',c.fim)||':'||
23 date_part('minute',c.fim)||':'||date_part('second',c.fim) as time)
24 ))
25 and b.ID_EVENTO = 'REVEILLON2014'
26 and a.LAT_LNG is null and b.LAT_LNG is null
27 group by ID, a.LAT_LNG;

```

SQL 4.5.e

Entrada: Local de residência presumida dos usuários, base de CDR, base de torres de telefonia.

Parâmetro: Torres de telefonia que atenderam o evento.

Saida: Quantidade de visitantes em eventos, assim como as regiões de origem inferida.

Quantidade de visitantes em eventos

```

1 select
2 a.dia as "DATA",
3 grao as cod_ibge,
4 nm_grao as "LOCAL",
5 nm_municip as "MUNICÍPIO",
6 zona as "ZONA",
7 pop_censo as "POPULAÇÃO",
8 pop_usu as "USUÁRIOS",
9 substring(

```

```

10     centroide,1,(strpos(centroide, ' ') -1)
11     ) as "LONGITUDE",
12     substring(
13     centroide,(strpos(centroide, ' ') +1)
14     ) as "LATITUDE",
15     praia as "DESTINO",
16     substring(
17     centroide_ev,1,(strpos(centroide_ev, ',') -1)
18     ) as "LAT_PRAIA",
19     substring(
20     centroide_ev,(strpos(centroide_ev, ',') +1)
21     ) as "LONG_PRAIA",
22     Distance_meters as "DISTÂNCIA(m)",
23     a.TX_DETEC as "TAXA",
24     a.Total_detectado as "TOTAL",
25     a.usuarios_in As "IN",
26     ((cast(a.usuarios_in as double precision))
27     /cast(a.Total_detectado as double precision)) as "PROB",
28     (((cast(a.usuarios_in as double precision))
29     /cast(a.Total_detectado as double precision))
30     * pop_censo) as "VISITANTES"
31 from
32     pop_usu_resultado_cruz a
33     inner join "Area21Grao_v2w_ev" b
34     on a.local=b.grao
35 order by
36     nm_grao;

```

A.3 Guia de tecnologias usadas

Uma breve descrição das tecnologias adotadas nesta tese será apresentada a seguir.

A.3.1 Processamento de dados

Para tornar possível que os dados de CDR fossem geo-referenciados em latitude e longitude, foi necessário processar o cadastro de todas as torres de telefonia da operadora que originou os dados, fazendo uma junção da chamada com a antena que atendeu-a. O ponto de atenção foi que geralmente as antenas ficam em trio na mesma posição, uma a 120 graus da outra para assegurar a área de cobertura ao redor. Ou seja, geralmente existe mais de uma antena capaz de processar chamadas na mesma latitude e longitude.

Foi usado o banco de dados colunar HP Vertica Community para processar os dados. A versão gratuita para pesquisa permite que até 1TB de dados sejam carregados. Não há limite de carga, função ou processamento. A tecnologia adotada no HP Vertica é baseada no conceito de banco de dados colunar (C-Store), ver LAMB *et al.* [76], que foi concebida especialmente para análise de dados, levando-se em conta que os modelos tradicionais não eram adequados para esse fim. O líder de pesquisa por trás do C-Store é o mesmo do PostgreSQL, Michael Stonebraker¹, ver STONEBRAKER *et al.* [77], do laboratório de pesquisas de ciência da computação e inteligência artificial do MIT.

A título de capacidade computacional, foi usado para processamento um servidor virtual com 12 CPUs Intel(R) Xeon(R) X5650 @ 2.67GHz, com cache de 12MB e memória de 16GB. O processamento mais lento foi o cálculo de matriz O-D por dia para todo o ano, que levou 3 dias para concluir.

¹<https://www.csail.mit.edu/user/1547>, visto em abril de 2017.

A.3.2 Visualização

A visualização de dados serve como ferramenta de apoio à estatística. Tomando o contexto de BIG DATA, é essencial possuir uma forma clara de se mostrar grandes conjuntos de dados. Sem um ferramental adequado de visualização torna-se praticamente impossível de se explorar e compreender os padrões apresentados pelos dados. Segundo TUKEY e TUKEY [78]:

"Não há nada melhor do que uma figura para você pensar em questões que tinha esquecido de perguntar (mesmo que mentalmente)."

Sem uma visualização adequada não há de se ter dúvida a respeito de padrões, pois ainda nem foi possível de se compreender o que os dados têm a dizer. Algumas ferramentas modernas de visualização para grandes conjuntos de dados podem ser vistas em UNWIN *et al.* [79] sob um enfoque científico. Num sentido mais jornalístico, YAU [80], explica que os dados são expressões da sociedade. Cada imagem é uma síntese de algo que vai embasar decisões ou simplesmente alavancar a compreensão de algum fenômeno por parte de uma audiência. Daí a importância de se criar uma imagem clara e direta ao público a que se destina.

Os dados de telefonia móvel chegam facilmente aos bilhões de registros. Nesse sentido, alguns estudos se propuseram a desenvolver técnicas de visualização adequadas a esse contexto. A Nokia lançou um desafio de dados (*Nokia Data Challenge*[81]), que consistia em disponibilizar uma grande quantidade de dados de telefonia para fomentar a realização de estudos nessa área. Um dos trabalhos apresentados foi o de SLINGSBY *et al.* [82], que permitiu explorar os aspectos espaciais e temporais de 38 anônimos durante 18 meses de forma interativa. Foram usados algoritmos já desenvolvidos em trabalhos anteriores [83] e [84]. O protótipo de visualização foi montado juntamente com a análise dos dados. Ou seja, a visualização foi tornando-se útil e acurada na medida em que os dados foram sendo compreendidos e novas perguntas foram sendo feitas.

Várias ferramentas de visualização atendem a demanda por visualização de grandes conjuntos de dados. Nesta pesquisa foram usadas três ferramentas: QGIS, Tableau² e Mapbox³.

O QGIS é um software livre voltado para processamentos e visualização de mapas. Ele suporta mapas vetoriais e raster (imagem). Possui diversos algoritmos prontos para realizar operações complexas em mapas, tais como: cálculo de polígonos de voronoi, checagem de topologia, interseção de pontos no polígono etc. Além de disponibilizar uma interface de "plugin" sob demanda que permite usufruir de pacotes desenvolvidos por usuários avançados, indo muito além do básico. Vide figura A.1 um exemplo de figura gerada pelo QGIS.

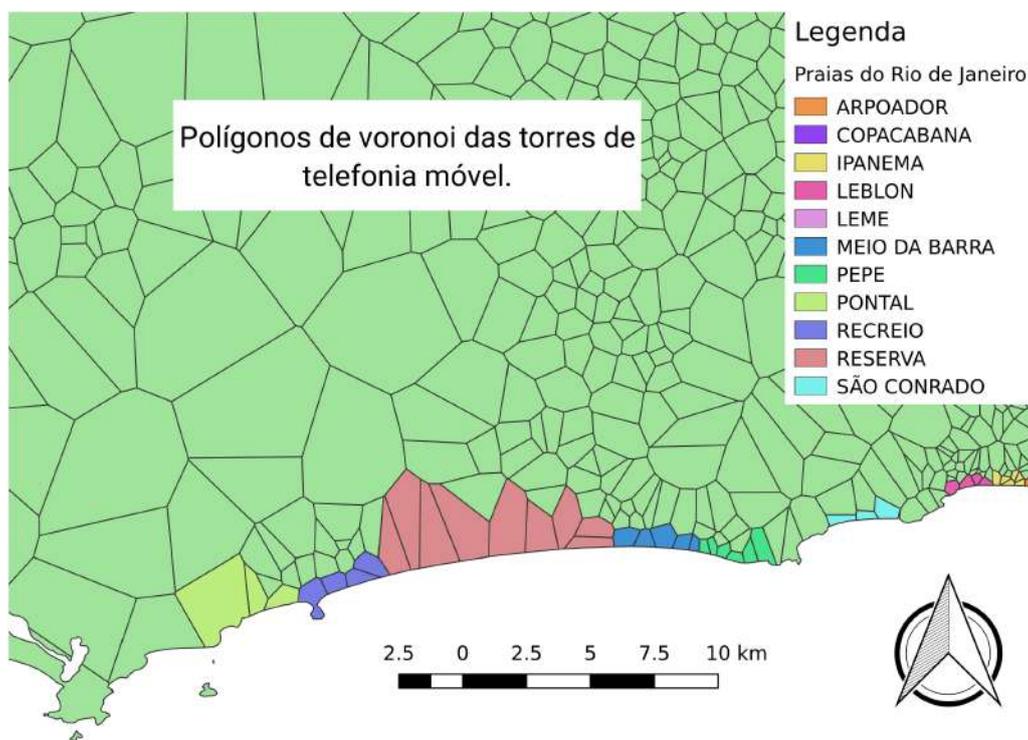


Figura A.1: Figura gerada com o software QGIS.

O QGIS também permite a integração com alguns bancos de dados especificamente desenvolvidos para lidar com mapas; como é o caso do PostGIS, um pacote de funções de mapa acoplados ao Postgresql. Essas funções se mostram

²Sob licença gratuita de 1 ano para estudante.

³<https://www.mapbox.com/>

muito úteis em conversão de sistema de coordenadas⁴, em cálculo de distância linear entre dois pontos, em junções de atributos por localidade espaciais etc.

As limitações do QGIS se apresentam de forma semelhante a muitos software livres; ocorrem quando a versão de algum dos pacotes que são pré-requisito de funcionamento não estão exatamente conforme o recomendado. Algumas funções disponibilizadas nos menus simplesmente não funcionam ou travam o sistema. Às vezes as soluções são fáceis a partir de pequenas atualizações. O mais seguro é não usar as versões "beta".

O Tableau é um pacote de visualização relativamente caro, contudo tem as suas compensações, assim como limitações. Trata-se de um pacote extremamente amplo, sendo a parte de mapas apenas uma das funcionalidades disponíveis. O foco do Tableau é um conceito: "tornar linhas e colunas compreensíveis às pessoas comuns, às pessoas que não são especialistas em análise de dados". Esse conceito vem praticamente mudando o mercado em torno da visualização de dados e em torno dos sistemas de BI tradicional. As pessoas continuam exercendo as próprias profissões e não precisam pedir que um especialista monte uma visualização, elas mesmo montam. Vide figura A.2 um exemplo de visualização disponibilizado para download no próprio repositório público da Tableau.

O diferencial do Tableau é a capacidade de conectar em quase todos os bancos de dados existentes. Contudo nem tudo é tão simples assim, a montagem de visualizações como a da figura A.2 exigem um nível elevado de conhecimento do produto.

O mapbox é uma plataforma on-line avançada de desenvolvimento de serviços baseados em mapas. No momento em que o desenvolver publica um mapa personalizado, ele já fica disponível on-line, e conseqüentemente como serviço disponível para ser integrado em outras plataformas compatíveis, como é o exemplo do próprio Tableau, que permite inserir um mapa personalizado do mapbox

⁴Latitude, longitude.

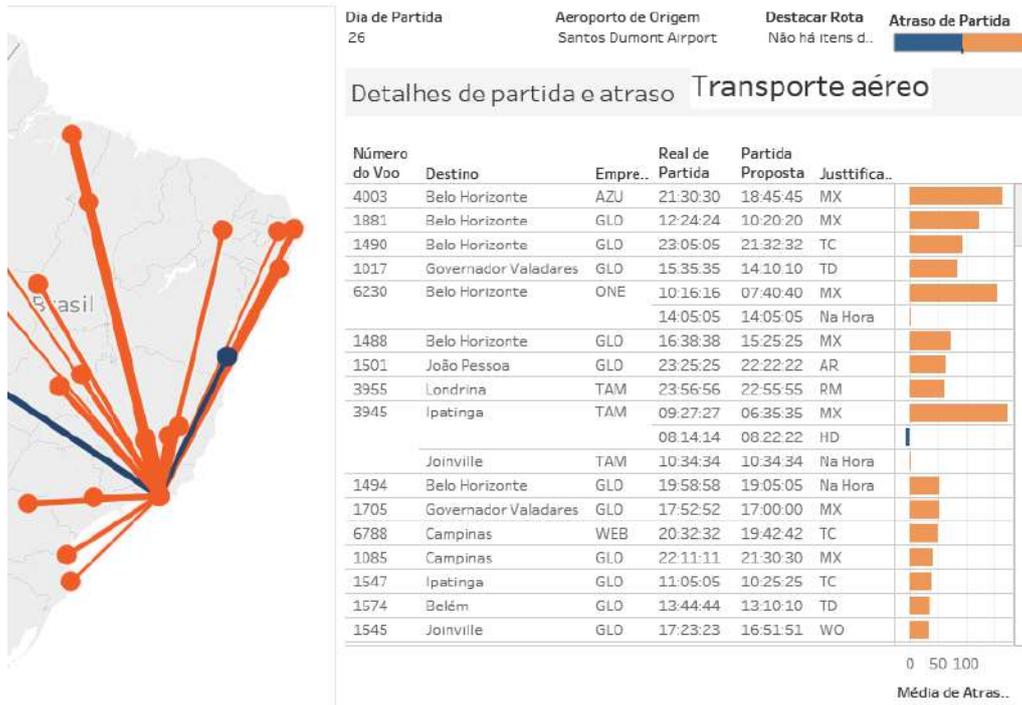


Figura A.2: Figura gerada com o software Tableau. Obtida a partir do site de exemplos públicos da própria Tableau.

dentro de uma visualização. Até 50 mil visualizações por mês, o serviço é gratuito. Vide figura A.3 um exemplo de mapa gerado pelo mapbox.



Figura A.3: Mapa gerado pelo mapbox.

Observe na figura A.3 que é dada ênfase nas vias expressas através da cor laranja e o recorte dos polígonos representa os subdistritos. Tudo o que é visto no mapa é personalizável, desde a cor do oceano até o tipo, cor e tamanho da fonte usado em certo nível de zoom, para nomear as cidades, por exemplo. Também é possível carregar o seu próprio shape file no mapbox, a versão gratuita permite que se carregue até 256MB.