



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TRANSPORTES**

**SAMEQUE FARIAS CUNHA DE OLIVEIRA**

**AS RELAÇÕES ENTRE MOBILIDADE URBANA, USO DO SOLO E  
DESLOCAMENTOS DE USUÁRIOS DE REDES SOCIAIS COMO APOIO AO  
PLANEJAMENTO DE TRANSPORTES**

**FORTALEZA**

**2020**

SAMEQUE FARIAS CUNHA DE OLIVEIRA

**AS RELAÇÕES ENTRE MOBILIDADE URBANA, USO DO SOLO E  
DESLOCAMENTOS DE USUÁRIOS DE REDES SOCIAIS COMO APOIO AO  
PLANEJAMENTO DE TRANSPORTES**

Tese apresentada ao Programa de Pós-Graduação em Engenharia, de Transportes da Universidade Federal do Ceará, como parte dos requisitos para a obtenção do título de Doutor em Engenharia de Transportes.

Orientador: Prof. Dr. Carlos Augusto Uchôa Da Silva

FORTALEZA

2020

SAMEQUE FARIAS CUNHA DE OLIVEIRA

**AS RELAÇÕES ENTRE MOBILIDADE URBANA, USO DO SOLO E  
DESLOCAMENTOS DE USUÁRIOS DE REDES SOCIAIS COMO APOIO AO  
PLANEJAMENTO DE TRANSPORTES**

Tese apresentada ao Programa de Pós-Graduação em Engenharia, de Transportes da Universidade Federal do Ceará, como parte dos requisitos para a obtenção do título de Doutor em Engenharia de Transportes.

Aprovada em: \_\_\_ / \_\_\_ / \_\_\_\_\_

BANCA EXAMINADORA

---

Prof. Dr. Carlos Augusto Uchôa Da Silva (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Bruno Vieira Bertoncini (Examinador interno)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Mário Angelo Nunes De Azevedo Filho (Examinador interno)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. André Luiz Cunha (Examinador externo)  
Universidade de São Paulo (EESC-USP)

---

Profa. Dra. Cira Souza Pitombo (Examinadora externa)  
Universidade de São Paulo (EESC-USP)

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

O51r Oliveira, Sameque Farias Cunha de.

As relações entre mobilidade urbana, uso do solo e deslocamentos de usuários de redes sociais como apoio ao planejamento de transportes / Sameque Farias Cunha de Oliveira. – 2021.  
121 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Transportes, Fortaleza, 2021.

Orientação: Prof. Dr. Carlos Augusto Uchôa Da Silva .

1. redes sociais. 2. padrão de deslocamento. 3. variabilidade de viagens. 4. similaridade entre matrizes.  
I. Título.

CDD 388

---

## AGRADECIMENTOS

Agradeço, em primeiro lugar, a Deus por o dom da vida, pela fé, pela perseverança e por toda a motivação para vencer todos os obstáculos que foram enfrentados nesse período. O antagonismo do doutorado é que escrever uma tese parece ser um caminho solitário, só que não é bem assim, pois todo o trabalho possui uma soma do apoio, companhia e torcida de familiares, amigos e colegas.

Aos meus pais Herotildes Farias e Marluce Maria, pela minha vida, pelo exemplo de dedicação, pela compreensão, pelo amor, pelo carinho, pelo apoio e por tudo que me ensinaram ao longo de todos esses anos, mostrando o verdadeiro caminho da vida.

Ao meu orientador Carlos Augusto Uchôa da Silva, por toda a paciência todos esses anos, por todas as alegrias, por todas as conquistas e por ter superado tantos obstáculos ao meu lado durante a realização desse trabalho. Em especial ao professor Mário Azevedo, que tenho como meu coorientador por toda a ajuda que me deu.

Aos meus irmãos Samaya Cristina e João Paulo Farias, por terem superado os reflexos dos meus momentos de angústia. Sou grato por o apoio incondicional de vocês.

As minhas primas Virgínia e Candice, por me inspirar através de gestos e palavras a superar os inúmeros problemas que surgiram durante todo esse percurso.

Aos professores Suelly Barroso, Verônica Castelo, Felipe Loureiro, Francisco Moraes, Bruno Bertocini, Heber e Manoel Mendonça, por acreditarem na minha capacidade de superar e vencer todos os obstáculos.

Aos meus colegas professores do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Aparecida, Lucicléia, Diego, Jandean, Carolina, Danielle Queiroz, Elizabete e Aluísio, por todo o apoio no final dessa jornada.

Aos meus amigos Edvan Moreira, Halini Magalhães, Jorge Azevedo, Emmeline Dantas, Luciano, Ana Lúcia e Berguly Oliveira, por todos os momentos de conversas, conselhos e mensagens de otimismo.

Ao meu amigo Bruno Tiago Angelo da Silva, por ter me apresentado o PETRAN, local esse que me proporcionou um crescimento vasto intelectual e profissional.

Aos meus amigos Rômulo Jácome, Annie, Zacarias, Clélia e Suyanne, com quem eu partilhava a vontade de cursar o doutorado. Agradeço por toda a influência e apoio para que eu tentasse uma vaga no programa.

Aos meus amigos de turma Juliana Brito, Cecília Lima, Macedo Filho, Lucas Cavalcante, Jorge Lucas, Regilene, Alessandro, Franco e Larissa, por todos os momentos compartilhados. E em especial, ao meu amigo Demóstenes Cassiano, por ter me ajudado a superar diversas dificuldades, por o conhecimento compartilhado e por toda a disponibilidade em me ajudar.

Aos amigos Daianne Fernandes, Ana Tália, Helry Luvillany, Patrícia, por dividir momentos de alegrias e tristezas. E em especial a Lisel Expósito, com quem tive o privilégio de conviver até o seu regresso à terra Natal.

Aos amigos do grupo Seridó's Garden, Lara Braide, Sasaki, Nara Gabriela, Joana Maia, Julie Anne, Davi Garcia, Kauê Braga, Edilany Aguiar, João Lucas, Leonardo, Diego, Arthur, Talyson, Carla Marília e Wendy. Obrigado por todos os momentos de descontração, apoio e dedicação. Agradeço, em especial, ao meu grande amigo Moisés Holanda, com quem partilhei bastante tempo nesses últimos anos, sem contar, com toda a ajuda nas correções de textos.

As minhas amigas Elliania Soares e Dona Maria, que com seus sorrisos e palavras faziam cada instante abençoado.

A todos os professores do Programa de Pós-Graduação em Engenharia de Transportes (PETRAN) que colaboraram e construíram bases sólidas no meu desenvolvimento e aprendizagem para o crescimento profissional. Seus nomes são inesquecíveis e por isso, dedico-lhes minha profunda admiração e respeito.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Finalmente, agradeço a Deus por me reservar saúde e proteção, e por me cercar de tantas pessoas maravilhosas.

Aos meus pais, Herotildes Farias e  
Marluce Maria.

## RESUMO

As Redes Sociais Baseadas em Localização (RSBL) permitem que os usuários confirmem as suas localizações através dos *check-ins*. Estudos recentes apontam que esses dados podem ser coletados e empregados em diversos estudos, como a obtenção de uma matriz OD. Métodos tradicionais de obtenção da matriz OD tem pouca atualização, ocorrendo a cada 10 anos. Na cidade de Fortaleza, por exemplo, ainda se usam dados coletados em 1996, através da pesquisa OD domiciliar. Assim, esta investigação teve como objetivo analisar a aplicabilidade dos dados provenientes de redes sociais como o *Twitter* e o *Instagram* para a obtenção de uma matriz OD. Com essa finalidade, sistematizou-se um método de coleta e tratamento de dados de *check-ins*, obtenção da matriz OD e definição dos motivos de viagens, além da identificação do perfil socioeconômico dos usuários. Consequente, efetuou-se uma análise de variabilidade das viagens entre os dias da semana e finais de semana. E por fim, realizou-se através de uma análise comparativa a relação entre a matriz OD das redes sociais e uma matriz OD externa. Os resultados indicaram que a maioria das viagens obtidas são por motivo Trabalho. Além disso, houve forte relação das origens (domicílios) de viagens com domicílios pertencentes a média e a alta renda, e uma baixa relação com indivíduos de baixa renda. As viagens obtidas a partir das redes sociais possuem uma espécie de transição entre os dias úteis e os finais de semana, mais precisamente às quintas-feiras, e como padrão para comprovar a confiabilidade dos dados, pode-se verificar que a frequência de deslocamentos por motivo Trabalho durante a semana é maior, e o contrário ocorre nos finais de semana. E com base em testes estatísticos, verificou-se a existência de similaridade entre as matrizes, com exceção dos destinos de viagens pelo motivo Outro, ou seja, a matriz OD das redes sociais referente a este motivo se comportou de modo diferente, quando comparada com a matriz OD do TRANUS pelo motivo Outro. De modo geral, essas matrizes das redes sociais podem ser aplicadas na área do planejamento de transportes, mais precisamente de complemento a pesquisa OD domiciliar. Além disso, ela pode servir de entrada para o modelo de distribuição, para que o fluxo de viagens dentro do meio urbano possa ser melhor compreendido. Embora essa matriz OD possa ser empregada em diversas áreas do planejamento de transportes, seus vieses e limitações devem ser enfatizados.

Palavras-chaves: redes sociais, padrão de deslocamento, variabilidade de viagens, similaridade entre matrizes



## ABSTRACT

Location-Based Social Networks (LBSN) allow users to confirm their locations through check-ins. Recent studies indicate that these data can be collected and employed in several studies, for instance, obtaining an OD matrix. Data obtained by traditional methods of obtaining the OD matrix are poorly updated, occurring every 10 years. In the city of Fortaleza, for example, data collected in 1996 through a household OD survey are still used. Thus, this research aimed to analyze the applicability of data from social networks, for instance, Twitter and Instagram, to obtain an OD matrix. In order to achieve this purpose, a method for collecting and processing data from check-ins, obtaining the OD matrix and defining the purposes of travel, in addition to identifying the socioeconomic profile of users was systematized. Subsequently, an analysis of travel variability between weekdays and weekends was carried out. Finally, the relationship between the OD matrix based on social networks data and an external OD matrix was established through a comparative analysis. The results indicated that most of the trips obtained are for work purposes. In addition, it was observed a strong relationship between the origins (households) of travel and households belonging to individuals of medium and high income, and a weak relationship with low-income individuals. The trips obtained from social networks data presented a sort of transition between working days and weekends, more precisely on Thursdays. In order to prove the reliability of the data, it can be verified that the frequency of travel for the purpose of work during the week is higher than that of the other purposes, and the opposite occurs on weekends. Based on statistical tests a similarity between the matrices was identified, except for travel destinations for the purpose Other, i.e., the OD matrix from social networks data referring to the purpose Other behaved differently when compared to the TRANUS OD matrix for this purpose. In general, these matrices of social networks data can be applied in the area of transport planning, more precisely as a complement to the household OD survey. Furthermore, it can be useful as an input to the distribution model, which can provide a better comprehension of the flow of travel within the urban environment. Although this OD matrix can be used in several areas of transport planning, its bias and limitations must be emphasized.

Keywords: social networks, displacement pattern, travel variability, similarity between matrices

## LISTA DE FIGURAS

Figura 1 – Limites dos bairros de Fortaleza .....	28
Figura 2 – Zonas de Tráfego.....	29
Figura 3 - Etapas metodológicas .....	30
Figura 4 - Raio de varredura (2 km) utilizado para a coleta de dados do <i>Twitter</i> , em projeção cartográfica UTM, fuso UTM 24S, datum SIRGAS 2000 .....	31
Figura 5 - Raio de varredura (5 km) utilizado para a coleta de dados do <i>Instagram</i> , em projeção cartográfica UTM, fuso UTM 24S, datum SIRGAS 2000 .....	32
Figura 6 – Mapa de lotes desagregados do município de Fortaleza-Sistema de atividades .....	35
Figura 7 - Algoritmo de classificação dos motivos de viagem, com base no tipo de lote de origem e do destino.....	36
Figura 8 – Exemplo de informações coletadas através do <i>Twitter</i> .....	39
Figura 9 - Informações coletadas através do <i>Instagram</i> .....	40
Figura 10 - Classificação a partir do lote de origem e destino .....	41
Figura 11 - Distribuição espacial das origens das viagens (base residencial).....	42
Figura 12 - Distribuição espacial das atividades do tipo Trabalho e Educação .....	43
Figura 13 - Distribuição espacial das atividades pelo motivo Outro.....	44
Figura 14 - Mapa de densidade dos domicílios componentes da amostra.....	45
Figura 15 - Correlação espacial dos domicílios por faixa de renda.....	46
Figura 16 - Dias da semana em função de viagens ao longo dos meses de análise .....	58
Figura 17 - Fluxograma para identificação do teste apropriado .....	60
Figura 18 - <i>Box plots</i> do número de viagens por usuário por dia da semana em Fortaleza – a) pelo motivo Trabalho e b) pelo motivo Outro .....	62
Figura 19 - Representação gráfica do teste Tukey (pelo motivo Trabalho) .....	66
Figura 20 - Representação gráfica do teste Tukey (pelo motivo Outro) .....	67
Figura 21 – Etapas Metodológicas .....	82
Figura 22 – Comparação entre <i>LISA Cluster Map</i> e gráfico de dispersão do modelo das origens de viagens das redes sociais (a) e o <i>LISA Cluster Map</i> e gráfico de dispersão do modelo das origens de viagens do TRANUS (b) por motivo Trabalho.....	86
Figura 23 - Comparação entre <i>LISA Cluster Map</i> e gráfico de dispersão do modelo das origens de viagens das redes sociais (a) e o <i>LISA Cluster Map</i> e gráfico de	

dispersão do modelo das origens de viagens do TRANUS (b) por motivo  
Outro ..... 87

Figura 24 - Comparação entre *LISA Cluster Map* e gráfico de dispersão do modelo das  
destino de viagens das redes sociais (a) e o *LISA Cluster Map* e gráfico de  
dispersão do modelo dos destinos de viagens do TRANUS (b) por motivo  
Trabalho..... 89

Figura 25 - Comparação entre *LISA Cluster Map* e gráfico de dispersão do modelo dos  
destinos de viagens das redes sociais (a) e o *LISA Cluster Map* e gráfico de  
dispersão do modelo dos destinos de viagens do TRANUS (b) por motivo  
Outro ..... 90

## LISTA DE TABELAS

Tabela 1 - Comparação dos Métodos emergentes versus tradicionais .....	19
Tabela 2 – Anova (pelo motivo Trabalho) .....	65
Tabela 3 – Anova (pelo motivo Outro) .....	65
Tabela 4 - Teste Tukey (pelo motivo Trabalho).....	66
Tabela 5 - Teste Tukey (pelo motivo Outro).....	67
Tabela 6 – Teste t.....	92
Tabela 7 – Teste t.....	92
Tabela 8 – Teste t.....	92
Tabela 9 – Teste t.....	92

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>16</b>
<b>1.1</b>	<b>Questões de pesquisa.....</b>	<b>20</b>
<b>1.2</b>	<b>Objetivos.....</b>	<b>20</b>
<b>1.3</b>	<b>Estrutura da pesquisa.....</b>	<b>21</b>
<b>2</b>	<b>CARACTERIZAÇÃO DE PADRÕES DE DESLOCAMENTOS URBANOS EM FORTALEZA COM A UTILIZAÇÃO DE DADOS DE REDES SOCIAIS GEORREFERENCIADAS .....</b>	<b>22</b>
<b>2.1</b>	<b>Introdução.....</b>	<b>22</b>
<b>2.2</b>	<b>Revisão bibliográfica.....</b>	<b>23</b>
<b>2.2.1</b>	<i>Estado da arte no uso de dados de redes sociais para estimação de padrões de deslocamentos.....</i>	<i>24</i>
<b>2.2.2</b>	<i>Considerações.....</i>	<i>26</i>
<b>2.3</b>	<b>Método .....</b>	<b>27</b>
<b>2.3.1</b>	<i>Descrição da área de estudo e zoneamento utilizados.....</i>	<i>27</i>
<b>2.3.2</b>	<i>Bases de dados utilizadas.....</i>	<i>29</i>
<b>2.3.2.1</b>	<i>Obtenção, tratamento e análise dos dados.....</i>	<i>30</i>
<b>2.3.2.2</b>	<i>Obtenção e leitura dos dados.....</i>	<i>30</i>
<b>2.3.2.3</b>	<i>Associação das postagens às ZATs.....</i>	<i>33</i>
<b>2.3.2.4</b>	<i>Exclusão das postagens que não se configuram deslocamentos .....</i>	<i>33</i>
<b>2.3.2.5</b>	<i>Classificação das postagens .....</i>	<i>33</i>
<b>2.3.2.6</b>	<i>Estimativa do local dos domicílios e identificação dos destinos das viagens .....</i>	<i>34</i>
<b>2.3.2.7</b>	<i>Atribuição do motivo da viagem .....</i>	<i>34</i>
<b>2.3.2.8</b>	<i>Obtenção das matrizes origem-destino amostrais.....</i>	<i>36</i>
<b>2.3.2.9</b>	<i>Caracterização socioeconômica dos domicílios.....</i>	<i>37</i>
<b>2.4</b>	<b>Resultados e discussões.....</b>	<b>38</b>
<b>2.4.1</b>	<i>Obtenção e leitura dos dados.....</i>	<i>38</i>
<b>2.4.2</b>	<i>Atribuição das viagens e construção das matrizes origem-destino .....</i>	<i>40</i>
<b>2.4.3</b>	<i>Caracterização socioeconômica dos domicílios.....</i>	<i>44</i>
<b>2.5</b>	<b>Considerações Finais.....</b>	<b>47</b>
<b>3</b>	<b>ANÁLISE DE VARIABILIDADE DOS DESLOCAMENTOS URBANOS A PARTIR DE DADOS DE REDES SOCIAIS EM FORTALEZA .....</b>	<b>49</b>
<b>3.1</b>	<b>Introdução.....</b>	<b>49</b>
<b>3.2</b>	<b>Revisão bibliográfica.....</b>	<b>50</b>
<b>3.2.1</b>	<i>Variabilidade comportamental dos usuários.....</i>	<i>50</i>
<b>3.2.2</b>	<i>Variabilidade de viagens através de métodos tradicionais .....</i>	<i>51</i>
<b>3.2.3</b>	<i>Variabilidade de viagens através de métodos alternativos.....</i>	<i>54</i>

3.2.4	<i>Considerações</i> .....	57
3.3	<b>Método</b> .....	57
3.3.1	<i>Tratamento preliminar dos dados</i> .....	59
3.3.2	<i>Caracterização dos padrões de viagens</i> .....	59
3.4	<b>Resultados e discussões</b> .....	61
3.4.1	<i>Caracterização dos padrões de viagens em Fortaleza</i> .....	62
3.4.2	<i>Testes estatísticos</i> .....	64
3.5	<b>Considerações finais</b> .....	69
4	<b>ANÁLISE COMPARATIVA ENTRE UMA MATRIZ OD OBTIDA ATRAVÉS DE USUÁRIOS DE REDES SOCIAIS E UMA MATRIZ OD EXTERNA E POSSÍVEIS APLICAÇÕES</b> .....	71
4.1	<b>Introdução</b> .....	71
4.2	<b>Revisão Bibliográfica</b> .....	72
4.2.1	<i>Matriz OD coletada a partir de dados das redes sociais</i> .....	72
4.2.2	<i>Potenciais e limitações dos dados advindos das redes sociais</i> .....	74
4.2.3	<i>Aplicabilidade de dados de redes sociais no planejamento de transporte</i> .....	76
4.2.3.1	<i>Detecção de Eventos</i> .....	77
4.2.3.2	<i>Mobilidade urbana e atividades</i> .....	78
4.2.4	<i>Considerações</i> .....	81
4.3	<b>Método</b> .....	81
4.3.1	<i>Obtenção das matrizes</i> .....	82
4.3.2	<i>Análise exploratória da comparação entre as matrizes</i> .....	83
4.3.3	<i>Teste de estatístico de comparação</i> .....	83
4.3.4	<i>Avaliação de aplicabilidade</i> .....	84
4.4	<b>Resultados</b> .....	84
4.4.1	<i>Análise exploratória</i> .....	85
4.4.1.1	<i>Número de viagens por motivo Trabalho originadas em uma dada Zona de tráfego</i> . 85	
4.4.1.2	<i>Número de viagens por motivo Outro, originadas em uma dada Zona de tráfego</i> ....	86
4.4.1.3	<i>Número de viagens por motivo Trabalho destinadas a uma dada zona de tráfego</i> ....	88
4.4.1.4	<i>Número de viagens por motivo Outro, destinadas a uma dada zona de tráfego</i> .....	89
4.4.2	<i>Análise comparativa</i> .....	91
4.4.3	<i>Avaliação de aplicabilidade das Matrizes OD de Redes Sociais</i> .....	93
4.5	<b>Considerações finais</b> .....	95
5	<b>CONCLUSÕES</b> .....	97
5.1	<b>Questões de pesquisa e conclusões</b> .....	97
5.2	<b>Sugestões para trabalhos futuros</b> .....	100
5.3	<b>Publicações no período</b> .....	100
	<b>REFERÊNCIAS</b> .....	102

**ANEXO 115**

<b>1: Rotina de coleta dos dados do <i>Twitter</i> .....</b>	<b>115</b>
<b>2: Rotina para estruturação, classificação e obtenção da matriz OD .....</b>	<b>118</b>

## 1 INTRODUÇÃO

Em planejamento e gestão das redes de transportes, a matriz origem e destino (OD) representa uma das mais importantes fontes de informações para o planejamento de transportes. Segundo Caceres *et al.* (2008), as matrizes OD são utilizadas para quantificar e sintetizar a mobilidade associada a pessoas ou bens, elas fornecem informações sobre o número de viagens realizadas entre pares OD durante um determinado intervalo de tempo. Dessa forma, fornece informações que representam a demanda de transportes em uma cidade.

Tradicionalmente, as áreas que compõem as origens e os destinos de uma matriz, são definidas por critérios socioeconômicos. Elas podem ser produzidas com diferentes níveis de agregação, dependendo do nível de detalhe desejado (zonas de tráfego, bairros ou setores censitários) ou do tipo de informação exigida: pelo modo de transporte ou pelo tipo de bem transportado (pessoas, mercadorias, etc.). Sua determinação é essencial e permite o aperfeiçoamento do uso do sistema de transportes, tanto para beneficiar os usuários em suas viagens diárias, quanto para se programar os investimentos necessários à adaptação de novas infraestruturas, caso sejam necessárias. A ausência deste planejamento indubitavelmente explica a maioria dos problemas como o congestionamento da rede observado em boa parte das grandes cidades.

A obtenção da matriz OD se dá através da pesquisa OD domiciliar, essa permite a coleta de informações detalhadas (WOLF *et al.*, 2000). Embora a pesquisa OD domiciliar proporcione uma amostra representativa da população, também possui algumas desvantagens em relação a sua obtenção, tais como: alto custo financeiro e atualização pouco frequente (HELLINGA, 1995). Dentre suas desvantagens, destaca-se a ausência de uma frequência recorrente de pesquisas desse tipo, esse fato em específico é considerado um dos mais graves. Isso provavelmente ocorre devido ao elevado custo para as prefeituras e a longa duração da realização desse tipo de pesquisa. Um exemplo desse problema a respeito da pesquisa OD domiciliar é visto na cidade de Fortaleza-Ce, cuja última pesquisa ocorreu em 1996. Uma outra desvantagem da pesquisa OD domiciliar refere-se à resistência por parte de alguns indivíduos em responder a pesquisa, desse modo, há uma necessidade de buscar métodos alternativos de forma a complementá-la. Essa complementação não vai substituir a pesquisa OD domiciliar, e sim, servir de artifício para cobrir a lacuna temporal existente.



Na pesquisa OD domiciliar de 1996, foram realizadas entrevistas em cinco mil domicílios, equivalente a 2% dos domicílios da cidade na época. Desde então, não houve mais pesquisas sobre os padrões de mobilidade na cidade. Apenas existiu um esforço de modelagem, tanto por parte da academia (SOUZA, 2016), quanto por parte da comunidade técnica, o que ocorreria em seus anos bases de análise (PREFEITURA MUNICIPAL DE FORTALEZA, 2015). Porém, ao levar em consideração o crescimento populacional em Fortaleza e o aumento das taxas de urbanização nos últimos anos, pode-se considerar que os dados da pesquisa OD domiciliar de 1996 estão defasados.

Em contrapartida, em 05 de fevereiro de 2019 foi iniciada uma pesquisa OD domiciliar, onde foram visitadas 23 mil residências, equivalente a aproximadamente 4% dos domicílios. Foram contratados cerca de 200 pesquisadores, que entrevistaram aproximadamente 100 mil pessoas (equivalente a 2% da população em 2019). O orçamento para a pesquisa que teve como objetivo principal obter os deslocamentos das famílias, ficou por volta de R\$ 11,3 milhões de reais. A data de divulgação estava prevista para janeiro de 2020, entretanto até julho de 2020, nenhum dado foi divulgado (OPOVO, 2019). Esse fato explicita a principal desvantagem desse tipo de método, que se refere à grande demanda de tempo para coleta, tabulação e divulgação dos resultados.

Recentemente, ocorreram mudanças que impactaram diretamente nos padrões de mobilidade da cidade de Fortaleza, com a implementação de novas infraestruturas de transportes, como o projeto de compartilhamento de bicicletas, construção de ciclofaixas e ciclovias, implementação de corredores de BRT, de trechos de VLT, de corredores exclusivos para transporte coletivo e ampliação da rede metroviária. Todas essas modificações tornaram imprescindível a obtenção de dados atualizados acerca da mobilidade e indicaram a real necessidade de uma nova pesquisa OD domiciliar ou do desenvolvimento de um método complementar para atualização dessas informações.

Em contrapartida, a ampla implantação de dispositivos de computação difundidos (*smartphone*, *smartcard*, aparelhos com tecnologia GPS, câmeras digitais, e outros) tem fornecido dados digitais sem precedentes, revelando detalhes tais como, onde estão localizados os usuários dessas tecnologias (espaço/tempo). Os avanços tecnológicos proporcionaram uma maior disseminação social de *smartphones*. À medida que essa tecnologia avançou, gerou o uso de diversos aplicativos, dentre eles, destacam-se as denominadas redes sociais. Essas, têm a capacidade de produzir uma grande

quantidade de dados por um custo ínfimo quando comparado aos métodos tradicionais (RASHIDI *et al.*, 2017). Atualmente o armazenamento desses dados na nuvem chega a gigabytes ao dia, ou seja, enquadra-se no chamado *Big data*. Esse termo é definido como uma base tão complexa e volumosa que se torna difícil de fazer algumas operações simples (remoção e ordenação) utilizando os recursos tradicionais disponíveis (VIEIRA *et al.*, 2012). Trabalhos anteriores têm utilizado diferentes metodologias para detectar a presença e o movimento desses usuários através de seus dados digitais (*Foursquare*, *Twitter* e *Flickr*) (QUERCIA *et al.*, 2011).

Dados de redes sociais é um dos métodos emergentes de coleta de dados no que diz respeito à compreensão da mobilidade nos últimos anos. A Figura 1 compara os prós e contras das fontes de dados de redes sociais com os métodos convencionais de coleta de dados, apresentando as características dos métodos de coleta de dados baseados em tecnologias recentes para a estimativa da compreensão da mobilidade (HU *et al.*, 2017). Métodos de coleta de dados incluindo pesquisa de viagens baseada em chamadas de telefones celulares (Yoo *et al.*, 2005; Deville *et al.*, 2014; Jarv *et al.*, 2014; Jiang *et al.*, 2017; Wang e Chen, 2018; Wu *et al.*, 2019; Forghani *et al.*, 2020). Método de coleta de dados a partir de Bluetooth (Versichele *et al.*, 2012; Michau *et al.*, 2014; Faye *et al.*, 2017; Yuan e Mills, 2019; Pu *et al.*, 2020), dados de redes sociais baseadas em localização como o *Foursquare* (Frith, 2013; Grabowicz *et al.*, 2014; Ribeiro *et al.*, 2014; Ferreira *et al.*, 2015; Abbasi *et al.*, 2017; Li *et al.*, 2018; Deeva *et al.*, 2019; Yang e Durarte, 2019; Chen *et al.*, 2020; Cerrone *et al.*, 2020; Novovic *et al.*, 2020) e dados do *Twitter* (Cranshaw *et al.*, 2012; Hong *et al.*, 2012; Li *et al.*, 2013; Hawelka *et al.*, 2014; Lenormand *et al.*, 2014; Llorente *et al.*, 2015; Jurdak *et al.*, 2015; Huang e Wong, 2015; Liu *et al.*, 2015; Kokkinogenis *et al.* 2015; Luo *et al.*, 2016; Yin *et al.*, 2016; Kurkcu *et al.*, 2016; Ellis *et al.*, 2016; Soliman *et al.*, 2017; Li *et al.*, 2017; Yin *et al.*, 2017; Khan *et al.*, 2017; Provenzano *et al.*, 2018; Wang e Taylor, 2018; Sulis *et al.* 2018; Kraemer *et al.*, 2018; Ahmouda *et al.*, 2019; Miyazawa *et al.*, 2019; Liao *et al.*, 2019; Bisanzio *et al.*, 2020; Martin *et al.*, 2020; Mazzoli *et al.*, 2020; Hu *et al.*, 2020; Yigitcanlar *et al.*, 2020; Luceri *et al.*, 2020; Ebrahimpour *et al.*, 2020; Paszto *et al.*, 2020).

Com base na Tabela 1, verifica-se que os métodos tradicionais, mais precisamente a pesquisa OD domiciliar é algo insubstituível e que as novas fontes de coletas de dados passivos podem servir de base complementar, ao passo que exista uma desatualização acerca das informações dos deslocamentos urbanos (referente a uma dada área). Apesar desses dados servirem de base para a compreensão dos deslocamentos

urbanos, esses possuem determinadas limitações. Uma mais frequentes refere-se à baixa quantidade de dados geolocalizados válidos para cada amostra (SLOAN e QUAN-HAASE, 2017; CERRONE *et al.*, 2015). Segundo Tasse e Hong (2017) existe uma baixa quantidade desses dados provenientes de áreas de baixa renda em comparação com o resto da cidade. Essa limitação se dá pela baixa imersão de *smartphones* e pelo acesso à internet em algumas áreas da cidade (TASSE e HONG, 2017). Uma outra limitação segundo Sloan e Quan-Haase (2017) refere-se ao fato de que uma parcela dos usuários do *Twitter* não ativa a função de geolocalização ao publicar *tweets*. Esse fato pode estar associado a fatores socioeconômicos (GRAHAM *et al.*, 2014).

Tabela 1 - Comparação dos Métodos emergentes versus tradicionais

Métodos de Coleta de Dados / Características	Métodos Trad. de Pesquisas	GPS	Bluetooth	Pesquisa com Smartphone	Sinais de Telefone Celular	Mídia Social	Redes Sociais
Resolução Espacial	Baixo	Baixo	Baixo	Alto	Alto	Alto	Alto
Resolução Temporal	Baixo	Alto	Alto	Alto	Alto	Alto	Alto
Implementação em Grande Escala	Sim	Não	Não	Não	Sim	Sim	Sim
Pesquisa/Custo dos Dados	Alto	Médio	Médio	Médio	Baixo	Baixo	Baixo
Necessidade da Pesquisa	Sim	Sim	Não	Sim	Não	Não	Não
Dados de Indicadores Sociais	Sim	Não	Não	Inferido	Não	Sim	Inferido
Dados de Origem/Destino	Sim	Sim	Sim	Sim	Sim	Sim	Inferido
Sequência da Viagem	Sim	Sim	Sim	Sim	Sim	Inferido	Inferido
Confirmação da Finalidade de Viagem	Sim	Limitado	Limitado	Sim	Limitado	Inferido	Sim
Modo de Distribuição	Sim	Inferido	Inferido	Sim	Inferido	Inferido	Inferido
Determinação do tempo de Chegada	Baixo	Alto	Alto	Alto	Alto	Alto	Alto
Determinação da Localização de Chegada	Baixo	Alto	Baixo	Baixo	Alto	Médio	Alto
Viés de Amostragem	Baixo	Médio	Médio	Médio	Médio	Sim	Sim
Privacidade	Não	Médio	Não	Não	Não	Médio	Médio

Autor: Adaptado de Hu (2017).

Por outro lado, embora um determinado conjunto de dados possam ser aceitáveis em termo de quantidades, a falta de representatividade foi reconhecida como um problema. Isso porque os dados das redes sociais coletados em locais específicos, revelam detalhes sobre a vida urbana nesses locais em particular, assim, um estudo referente a uma dada área, se limita a um determinado local, e suas conclusões podem ser diferentes de uma outra área de estudo (MARTI *et al.*, 2019). De outro modo, alguns autores (Agryzkov *et al.*, 2015 e Marti *et al.*, 2017) revelam que esses dados podem

representar toda a população. De modo geral, os dados advindos das redes sociais, possuem as vantagens de serem coletados em grande escala e em tempo real, permitindo dessa forma a incorporação de dinâmicas temporais na análise dos padrões de viagens dos indivíduos. As preocupações mencionadas são reconhecidas em estudos que utilizaram dados de redes sociais no que se refere a compreensão dos padrões de deslocamentos urbanos. No entanto, os desafios e limitações acerca da aplicação desses dados de modo a complementar a pesquisa OD domiciliar e a compreensão da representatividade continuam sendo um desafio.

### 1.1 Questões de pesquisa

Diante das lacunas encontradas na literatura em relação à aplicação dos dados de redes sociais para a caracterização do padrão de deslocamentos, bem como a variabilidade do comportamento de viagens e as possíveis aplicabilidade dessa matriz OD, levantam-se os seguintes questionamentos:

- *Como definir e caracterizar os deslocamentos a partir de dados geolocalizados advindos das redes sociais?*
- *Como as viagens baseadas em redes sociais variam ao longo dos diferentes dias da semana?*
- *Como investigar as semelhanças e diferenças entre a matriz OD das redes sociais e uma matriz OD externa?*

### 1.2 Objetivos

Diante do exposto, este trabalho propõe analisar a aplicabilidade dos dados provenientes de redes sociais como o *Twitter* e o *Instagram* para a obtenção de uma matriz OD na cidade Fortaleza-Ce. Para contemplar este objetivo principal, foram definidos os seguintes objetivos específicos:

- Desenvolver um método para definir e caracterizar o padrão de deslocamentos no meio urbano a partir de dados provenientes de redes sociais baseadas em localização;

- Investigar o grau de variação do comportamento de viagens dos usuários do sistema de transportes em Fortaleza utilizando dados de redes sociais;
- Investigar o nível de aderência das informações que possam ser extraídas da matriz OD advinda das redes sociais e uma determinada matriz OD externa.

### 1.3 Estrutura da pesquisa

CAPÍTULO 1: INTRODUÇÃO - Apresenta o contexto da investigação e a definição do problema de pesquisa, além dos objetivos geral e específicos definidos com a finalidade de preencher as lacunas encontradas a respeito da problemática exposta. Destaca-se que os Capítulos 2, 3, 4 e 5 serão apresentados em formato de artigo científico.

CAPÍTULO 2: CARACTERIZAÇÃO DE PADRÕES DE DESLOCAMENTOS URBANOS EM FORTALEZA COM A UTILIZAÇÃO DE DADOS DE REDES SOCIAIS GEORREFERENCIADAS – Descreve uma metodologia capaz de obter, analisar, definir e caracterizar origens e destinos de viagens a partir de dados provenientes do *Instagram* e do *Twitter*, além de compreender a que classe socioeconômica pertence os indivíduos que realizam essas viagens.

CAPÍTULO 3: ANÁLISE DE VARIABILIDADE DOS DESLOCAMENTOS URBANOS A PARTIR DE DADOS DE REDES SOCIAIS EM FORTALEZA – Dedicar-se à análise do fenômeno da variabilidade de viagens em Fortaleza, utilizando viagens advindas do *Instagram* e *Twitter*.

CAPÍTULO 4: ANÁLISE COMPARATIVA ENTRE UMA MATRIZ OD OBTIDA ATRAVÉS DE USUÁRIOS DE REDES SOCIAIS E UMA MATRIZ OD EXTERNA E POSSÍVEIS APLICAÇÕES - Apresenta uma análise comparativa e sistemática entre a matriz OD das redes sociais e uma matriz OD externa.

CAPÍTULO 5: CONCLUSÕES – Apresenta brevemente as conclusões e ponderações deste estudo, além de sugestões para trabalhos futuros.

## 2 CARACTERIZAÇÃO DE PADRÕES DE DESLOCAMENTOS URBANOS EM FORTALEZA COM A UTILIZAÇÃO DE DADOS DE REDES SOCIAIS GEORREFERENCIADAS

### 2.1 Introdução

O conhecimento dos padrões de deslocamentos de pessoas, em especial das origens e destinos de suas viagens, tem se mostrado essencial no processo de planejamento de transportes. Para a obtenção desse tipo de informação é preciso recorrer, convencionalmente, às pesquisas OD domiciliares, a partir das quais são obtidas informações acerca dos padrões de viagens dos indivíduos (motivo da viagem, origem e destino, modo utilizado, dentre outras), além de suas características socioeconômicas. O principal produto desse tipo de pesquisa é a estimação de uma matriz origem-destino, com todos os deslocamentos realizados ao longo de um dia típico.

Reconhece-se, entretanto, que tal tipo de abordagem apresenta algumas limitações, relativas tanto à operacionalização da pesquisa quanto à qualidade das informações obtidas. No primeiro caso, a principal limitação está relacionada aos elevados custos, tanto financeiros quanto de tempo (normalmente mais de um ano), para a execução desse tipo de pesquisa. Além disso, é comum verificar erros inerentes ao processo de coleta, tabulação e armazenagem dos dados (HELLINGA, 1995). Do ponto de vista da qualidade da informação obtida, o que mais se questiona nesse tipo de pesquisa é o fato dela retratar apenas um recorte temporal específico, normalmente um dia da semana considerado típico. Logo, a matriz obtida apresenta deficiências, já que não consegue incorporar as dinâmicas espaciais e temporais inerentes ao fenômeno analisado (IQBAL *et al.*, 2014).

Tais limitações reforçaram a necessidade de se buscar fontes alternativas de dados de viagens, de modo a preencher as lacunas observadas nas informações providas de pesquisas OD domiciliares (FERREIRA, 2013; WOLF *et al.*, 2003). Dentre algumas alternativas, se destaca a utilização de dados provenientes de redes sociais baseadas em localização (*RSBL*), como o *Twitter* e o *Instagram*, por exemplo. Estes tipos de aplicativos possibilitam a obtenção da localização geográfica dos usuários no momento em que estes compartilham suas atividades através dos chamados *check-ins*. Isto é possível pela incorporação de sistemas globais de posicionamento por satélites (*GNSS*) aos telefones

móveis. Tais informações apresentam a grande vantagem de serem coletadas em grande escala e em tempo real permitindo dessa forma a incorporação de dinâmicas temporais na análise do padrão de viagens dos indivíduos (HU e JIN, 2017). Portanto, o uso dessas ferramentas, quando combinadas a outras bases de dados (uso do solo, censo demográfico, por exemplo), despontam como potenciais fontes de informação para se estimar padrões de mobilidade no meio urbano, informação essencial ao processo de planejamento (AGRYZKOV *et al.*, 2017).

Diante desse contexto, o objetivo principal deste capítulo é desenvolver um método para definir e caracterizar o padrão de deslocamentos no meio urbano a partir de dados provenientes de redes sociais baseadas em localização. Para isso, primeiramente foi feita uma revisão da literatura (Seção 2.2), cujo objetivo foi a identificação de lacunas nessa abordagem a partir do levantamento de trabalhos que utilizaram dados de redes sociais para caracterização de padrões de mobilidade no meio urbano. A Seção 2.3 introduz o método utilizado para a análise, que vai desde a obtenção e tratamento dos dados, estimação dos motivos das viagens e construção das matrizes origem-destino, além da identificação do perfil socioeconômico dos domicílios amostrados. Já na Seção 2.4, é apresentada a aplicação do método proposto a partir de um estudo de caso para a cidade de Fortaleza. A partir dos resultados obtidos, foi feita uma discussão do método apresentado, analisando as principais limitações encontradas. Por fim, a Seção 2.5 traz as principais conclusões do trabalho.

## **2.2 Revisão bibliográfica**

Na década 2010-2020, houve um crescimento no interesse da comunidade científica em explorar a utilização de dados de redes sociais como instrumento de obtenção de padrões de deslocamentos de indivíduos no meio urbano. Os parágrafos a seguir trazem um resumo do estado da arte nesta temática, identificando suas limitações, além de lacunas ainda não exploradas pela literatura.

### ***2.2.1 Estado da arte no uso de dados de redes sociais para estimação de padrões de deslocamentos***

Um dos primeiros trabalhos com temática baseada na utilização de dados de redes sociais como ferramentas para a compreensão de características socioeconômicas de indivíduos foi o estudo de (CRANSHAW *et al.*, 2010). Eles propuseram um modelo de agrupamento para estudar a mobilidade da cidade de Pittsburgh com base nos dados gerados pelas redes sociais. A metodologia aplicada obteve aproximadamente 42787 *check-ins* do *Twitter*. Além disso, um algoritmo foi desenvolvido capaz de varrer uma determinada área e encontrar *clusters*, que posteriormente foram denominados de *lifelihoods* (representação de áreas dinâmicas da cidade). Com o intuito de validar essas áreas, os autores realizaram entrevistas de uma hora com habitantes da região de estudo. Os resultados indicaram que os *clusters* obtidos podem fornecer um bom suporte para o entendimento da dinâmica acerca das atividades realizadas pelos usuários dentro do meio urbano.

Cheng *et al.* (2011) utilizaram dados do *Twitter*, referentes a aproximadamente 22 milhões de *check-ins* de usuários do mundo inteiro. O principal objetivo da pesquisa foi a estimação de características socioeconômicas desses usuários a partir de aspectos espaciais e temporais das postagens. Foi verificado que a maior densidade de *check-ins* aconteceu em regiões da América do Norte, Europa Ocidental e Ásia Meridional e Pacífica. Além disso, foi percebida uma maior concentração de postagem às 9h, 12h e 18h. O perfil socioeconômico dos usuários foi determinado com base na sua influência na rede social. Desta forma, ficou evidente que indivíduos com grande quantidade de seguidores pertenciam a classes sociais mais elevadas.

Hasan *et al.* (2013) avançaram no uso desses tipos de dados, ao inferir padrões de atividades realizadas por indivíduos a partir de dados do *Twitter* e do *Foursquare*. Foram coletadas informações de *check-ins* dos usuários dessas redes ao longo de 10 meses, em três cidades americanas (Nova Iorque, Chicago e Los Angeles). Com esses dados, os autores construíram um mapa com a distribuição das atividades dos indivíduos da amostra, destacando os locais mais frequentados por esses usuários. Essa amostra foi, então, extrapolada com a utilização de métodos de estimação de densidade de *Kernel*, cujo produto foi a obtenção de mapas de atividades nas três cidades analisadas para períodos temporais distintos, ao longo do dia. Apesar de trazer avanços no conhecimento



do perfil socioeconômico e no padrão de atividades dos usuários de redes sociais, os trabalhos anteriormente citados, pouco contribuíram para a compreensão dos deslocamentos desses indivíduos no meio urbano. Para preencher essa lacuna, Gao *et al.* (2014) utilizaram dados provenientes do *Twitter* para estimar uma matriz origem-destino, no âmbito regional, com os deslocamentos entre cidades da Região Metropolitana da Califórnia, nos Estados Unidos. Para isso, foi coletada uma amostra de aproximadamente 6,8 milhões de *tweets* georreferenciados de mais de 110 mil usuários. O método utilizado para se determinar as viagens baseou-se em duas etapas: detecção do deslocamento do indivíduo e agregação da viagem na área. Na primeira etapa, a trajetória do indivíduo era identificada pela sequência de *tweets* geolocalizados. Uma vez definidos o início e o final desse deslocamento, cada viagem terá uma origem e um destino, no nível de agregação considerado.

Outro trabalho cuja proposta foi a estimação de uma matriz origem-destino a partir de dados de redes sociais foi o de (ABBASI *et al.*, 2017). Em sua pesquisa, eles aplicaram um método probabilístico de estimação de deslocamentos de indivíduos baseado em ranqueamento, através do qual é determinada a probabilidade de determinado destino dada uma origem. A principal dificuldade deste método, entretanto, está justamente em ranquear esses potenciais destinos. Deste modo, os autores aplicaram três métodos distintos de definição desse ranqueamento: um no qual o peso dos possíveis destinos era determinado pela distância destes à origem; um segundo baseado na quantidade de estabelecimentos entre a origem e o destino; e um terceiro que ponderava a quantidade de estabelecimentos pelo total de *check-ins* realizados nos mesmos. A análise foi realizada em Nova Iorque e utilizou dados provenientes do *Foursquare*. Os resultados dos três modelos foram comparados com os dados de viagens de táxi fornecidos pela Comissão de Táxis e *Limousines* (TLC). Concluíram que o uso dos dados da rede social melhorou a capacidade de previsão do modelo, em comparação aos dois primeiros métodos, devido à capacidade de se incorporar dinâmicas do comportamento humano através das informações de *check-ins*.

Molloy e Moeckel (2017) modelaram a atratividade de destino de viagem em Ontário através dos *check-ins* oriundos do *Foursquare*. A metodologia consistiu na agregação dos *check-ins* como indicadores de atratividade de destino para categorias específicas. Para isso, foi aplicado o modelo *logit multinomial*, e este estimou o potencial dos *check-ins*, em medir a atratividade. Os resultados mostraram que a precisão do

modelo em todos os fins de viagem melhorou significativamente, particularmente para viagens de lazer.

Cheng *et al.* (2018) desenvolveram um modelo de regressão *random forest* para estimar o total de viagens entre pares OD da região metropolitana de Sydney, na Austrália. Foram construídos dois modelos distintos: um com a utilização de dados de viagens de uma pesquisa OD domiciliar, além de informações do censo demográfico; e um segundo modelo incorporando dados de *check-ins* do *Twitter* em substituição à pesquisa OD domiciliar. Neste segundo modelo, utilizou-se uma amostra de mais de um milhão de *tweets* georreferenciados, referentes a mais de 170 mil usuários. Os autores obtiveram uma melhor precisão e estabilidade no modelo de regressão, ao considerar os dados do *Twitter* no modelo.

Por fim, Oliveira *et al.* (2019) obtiveram a posição dos usuários do *Twitter* e do *Instagram*, com o objetivo de identificar o uso do solo da cidade de Fortaleza-CE. O método consistiu na aplicação de um algoritmo e o uso do site *Netlytic* para aquisição dos dados, que foram mapeados e confrontados às informações sobre os diferentes tipos de uso do solo, fornecidas pela prefeitura municipal. Os resultados indicaram que o método apresentou uma alta correlação espacial entre o uso do solo inferido e o solo oficial, permitindo a identificação de áreas domiciliares e comerciais.

### 2.2.2 Considerações

A partir da revisão realizada foram percebidas algumas lacunas, principalmente de cunho metodológico, nesses trabalhos. A primeira delas foi a ausência de um método para a identificação dos domicílios dos indivíduos amostrados (GAO *et al.*, 2014; ABBASI *et al.*, 2017; CHENG *et al.*, 2018). Desta forma, não foi possível diferenciar viagens de base domiciliar daquelas de base não domiciliar. Além disso, nos trabalhos citados não foram apresentados métodos para especificar as atividades realizadas nos destinos, de modo a classificar as viagens por motivo. Foram observadas ainda limitações com relação à caracterização socioeconômica dos indivíduos amostrados. Cheng *et al.*, (2011), por exemplo, fizeram essa caracterização apenas com base no número de seguidores dos usuários, em que se assumiu que quanto maior o número de seguidores maior seria a renda do indivíduo.

Apesar de todo o esforço metodológico com a finalidade de suprir essas deficiências, o método proposto neste capítulo se baseou em bases de dados agregados e de períodos distintos das coletas de dados realizadas. Dessa forma, existe uma defasagem temporal no método proposto.

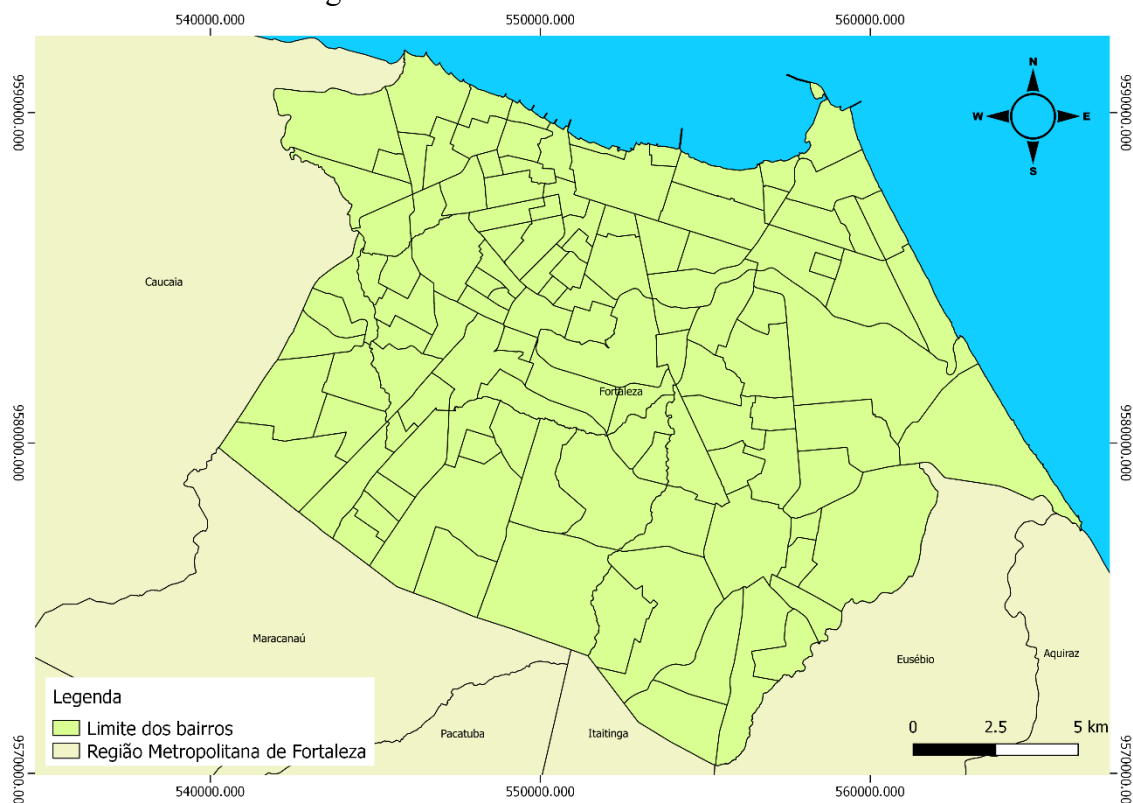
## **2.3 Método**

Nesta seção são descritos os passos metodológicos para se estimar padrões de deslocamentos na cidade de Fortaleza com o uso de dados de rede sociais. Para isso, primeiramente são descritos a área de estudo e o zoneamento utilizados. Em seguida, é apresentada a base de dados obtida, detalhando as etapas de coleta e tratamento dessas informações. Por fim, é introduzido o método utilizado para a construção de uma matriz origem-destino para a região, que passa pela identificação das origens e destinos das observações da amostra, dos motivos das viagens, além da caracterização socioeconômica dos indivíduos.

### ***2.3.1 Descrição da área de estudo e zoneamento utilizados***

O método aqui proposto foi aplicado na cidade de Fortaleza, localizada no litoral do estado do Ceará, nordeste brasileiro. Em termos populacionais, é a quinta maior cidade do Brasil, com 2,6 milhões de habitantes, os quais ocupam uma área de aproximadamente 315 km<sup>2</sup>. Tais características fazem com que a capital cearense possua a maior densidade demográfica dentre as capitais brasileiras, com quase 7800 hab/km<sup>2</sup>. Sob a ótica territorial, Fortaleza é composta por 121 bairros, sendo ela a principal cidade de uma região metropolitana composta por outros 18 municípios, os quais também impactam no sistema de transportes da cidade como apresenta a Figura 1.

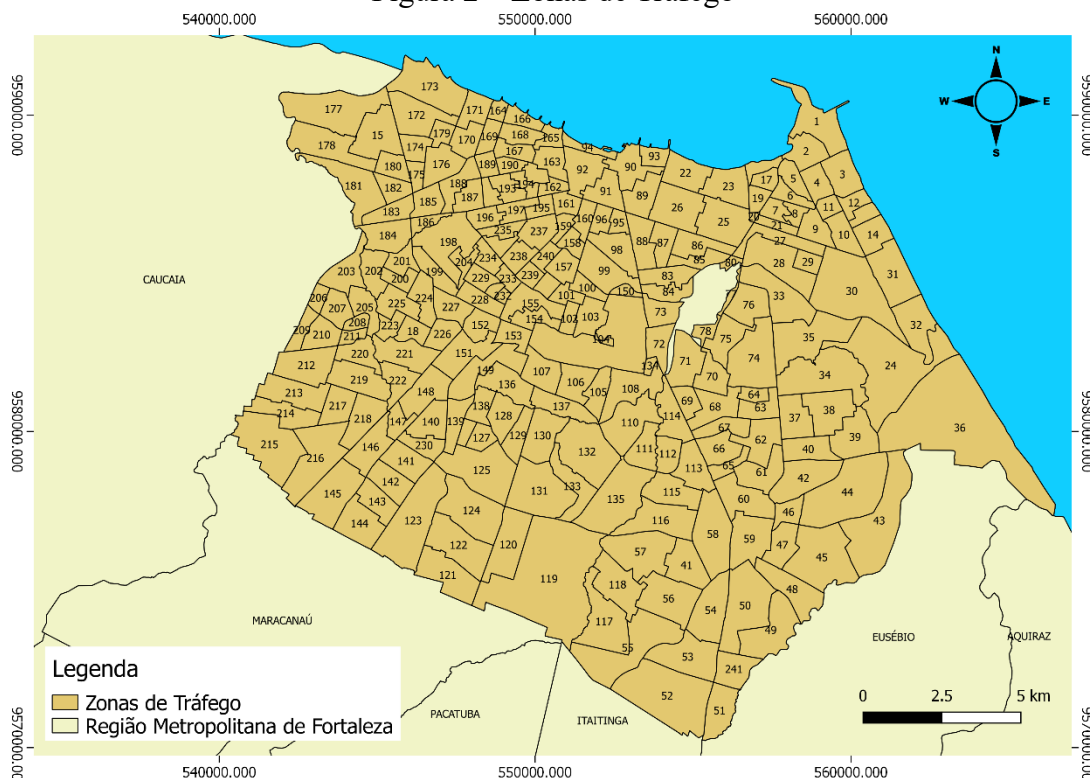
Figura 1 – Limites dos bairros de Fortaleza



Fonte: Elaborado pelo autor.

Diante de tal configuração, optou-se, neste trabalho, pelo uso do zoneamento proposto por LIMA (2017) como mostra a Figura 2. Neste zoneamento, a cidade de Fortaleza foi dividida em 241 ZATs (Zonas de Análise de Tráfego), formadas a partir da agregação dos setores censitários do censo do ano 2000, com base na homogeneidade socioeconômica. Onde cada zona possui um raio médio de 500 metros, de onde se infere que a maior parte das viagens intrazonais é realizada por modos não motorizados.

Figura 2 – Zonas de Tráfego



Fonte: Elaborado pelo autor.

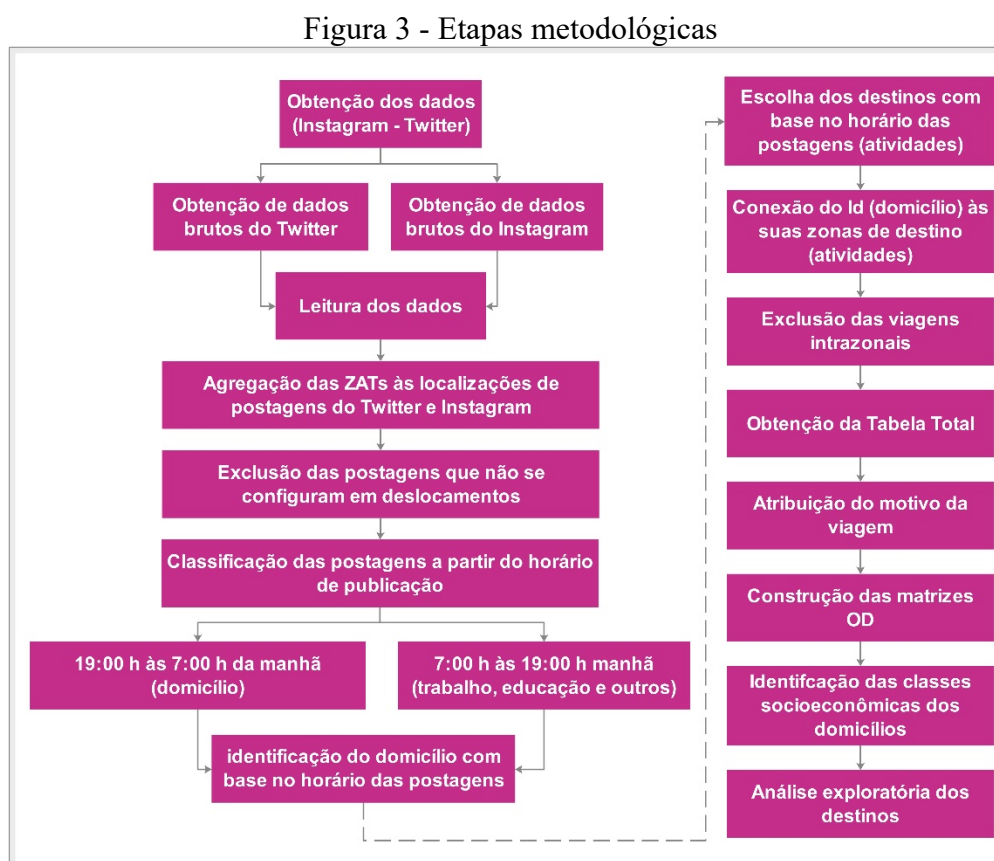
Vale destacar que, embora ciente da influência dos municípios da região metropolitana, estes foram desconsiderados no zoneamento proposto, devido à indisponibilidade de dados relativos a essas regiões.

### 2.3.2 Bases de dados utilizadas

Neste estudo, foram utilizadas duas bases de dados provenientes de RSBL: *Twitter* e *Instagram*. A primeira é considerada um *microblog*, no qual os usuários podem interagir através de mensagens de no máximo 280 caracteres, compartilhando, dentre outras informações, atividades realizadas e locais visitados (BIFET e FRANK, 2010). Já o *Instagram* tem como base o compartilhamento de fotos e vídeos, através dos quais é possível identificar as atividades realizadas e a localização dos usuários, através dos *check-ins*, ferramenta disponibilizada pelo aplicativo (PIZA, 2012). Além destes, foram também utilizados dados de uso do solo, fornecidos pela Secretaria de Finanças do município (SEFIN).

### 2.3.2.1 Obtenção, tratamento e análise dos dados

A Figura 3 apresenta o fluxograma das etapas do método, que engloba a obtenção e tratamento dos dados, estimação de matrizes origem-destino, por motivo, e inferência da classe econômica dos domicílios dos usuários.



Fonte: Elaborado pelo autor.

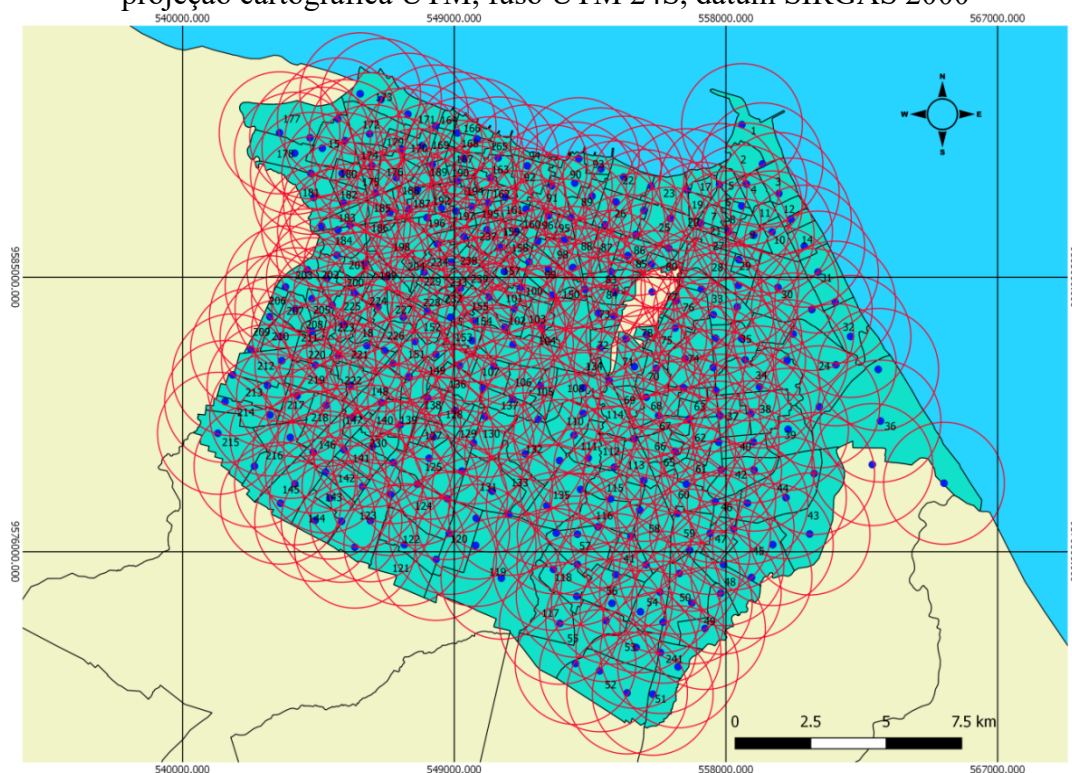
### 2.3.2.2 Obtenção e leitura dos dados

O processo de coleta dos dados do *Twitter* foi iniciado através do link a seguir: <https://apps.Twitter.com/>. Após o acesso da página, indicada anteriormente, foi necessário realizar um login e criar um projeto (app). Para criar um app foram preenchidas algumas informações, tais como: nome ao qual foi nomeado de *Dscan* e descrição do aplicativo. Logo após a criação do app foram obtidas duas chaves de acesso, sendo elas: *Consumer Key* e a *Consumer Secret*. Esses dois *tokens* deram acesso à *Application Programming Interface* - API. Para a coleta ocorrer foi fundamental criar uma rotina em

linguagem python e empregar a biblioteca chamada *twython* (ALZHRANI, 2018). Para a execução do *twython* se inseriu na rotina os tokens criados (Anexo 1). Após a implementação do *twython*, esse utiliza dois parâmetros, sendo eles: a posição (coordenadas), de onde se deseja coletar o *tweet* e o raio de varredura.

Para a obtenção dos dados em Fortaleza, aplicou-se um raio de varredura de 2 km, para cada um dos 241 centroides identificados através de coordenadas planas em projeção cartográfica UTM (N, E) das ZATs (Figura 4), de modo a abranger toda a área da cidade. Além disso, foram testados diversos valores para o raio de varredura, porém apenas o raio de 2 Km forneceu o resultado esperado, de outro modo, apenas o raio de 2 Km fornece o último *tweet* postado na zona. Nessa busca, foram coletados todos os *tweets* publicados, dentro do raio pré-definido, desde o momento do início da varredura até o momento em que se decidir interromper a rotina.

Figura 4 - Raio de varredura (2 km) utilizado para a coleta de dados do *Twitter*, em projeção cartográfica UTM, fuso UTM 24S, datum SIRGAS 2000



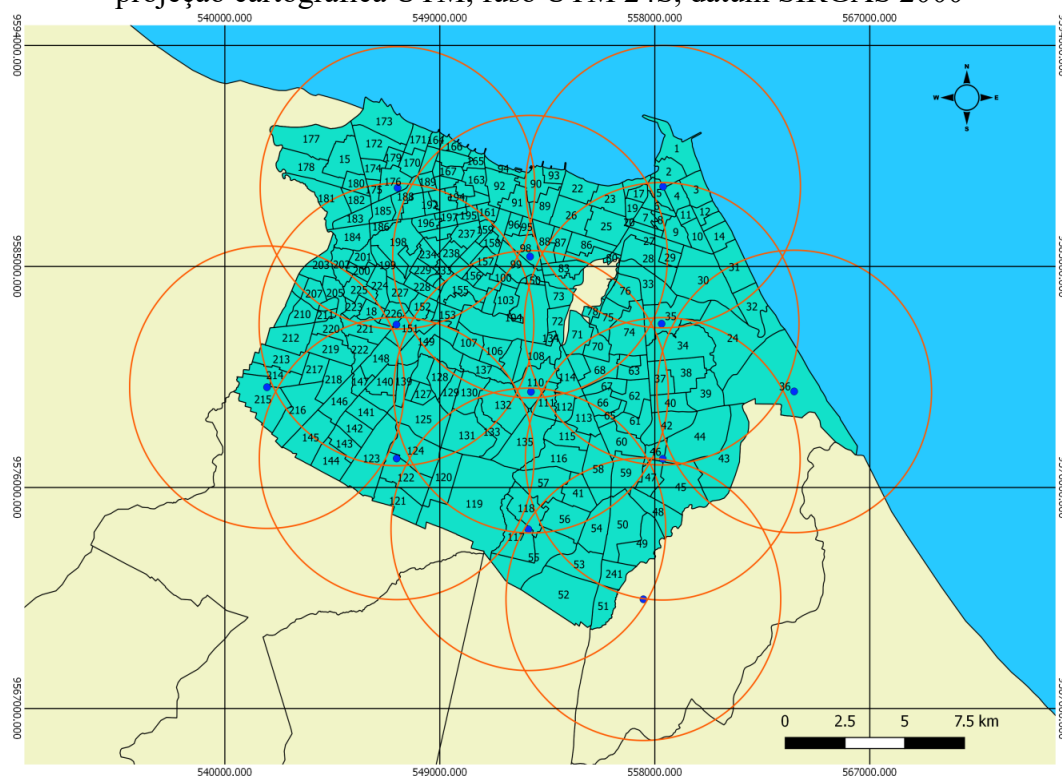
Fonte: Elaborado pelo autor.

Os dados provenientes do *Instagram* foram obtidos através do site <https://netlytic.org>. Em seguida, foram definidos os centroides e raio de varredura de 5km (predefinido pelo site) para coleta das informações. O site disponibiliza então os pontos

de coleta, totalizando 12 centroides para recobrir integralmente a área de estudo. A Figura 5 ilustra as áreas definidas para a coleta dos dados do *Instagram*.

O tempo de coleta dos dados foi realizado durante o intervalo de um ano, que foi do dia 01/10/2017 a 31/10/2018. Porém, a faixa de tempo utilizada neste trabalho é referente a oito meses (fevereiro a junho e de agosto a outubro de 2018). Além disso, os dados obtidos pelas duas redes são advindos de perfis públicos, dessa forma, não existiu a possibilidade de coleta de perfis privados.

Figura 5 - Raio de varredura (5 km) utilizado para a coleta de dados do *Instagram*, em projeção cartográfica UTM, fuso UTM 24S, datum SIRGAS 2000



Fonte: Elaborado pelo autor.

Os dados brutos obtidos das duas redes sociais foram, então, convertidos em formato de texto, \*.CSV. Dessa forma, foi possível filtrar, através de uma rotina implementada em linguagem R, apenas as informações necessárias para a análise: ID do usuário e data, horário e coordenadas geográficas da postagem. Após a filtragem, as bases de dados foram unificadas. Porém, para evitar erros de duplicação, ou seja, quando um mesmo usuário publica nas duas redes sociais ao mesmo tempo, foi tomado o devido cuidado em verificar se havia publicações realizadas em um mesmo horário (hora, minuto



e segundos), coordenadas iguais. Assim, caso houvesse uma postagem duplicada, seria considerada apenas uma única postagem.

### 2.3.2.3 Associação das postagens às ZATs

Após a leitura e filtragem dos dados, com o uso de uma rotina (Anexo 2) em linguagem R, foi realizada a associação dos pontos georreferenciados às respectivas ZATs, dentre as 241 existentes. Tem-se, portanto, em  $L_u = (l_u^1, l_u^2, \dots, l_u^n)$  a sequência dos locais ( $L$ ) dos *tweets* e fotos (*Instagram*) geolocalizados de um determinado usuário ( $u$ ). Esses pontos são, então, associados espacialmente às suas respectivas ZATs, formando, dessa maneira, um conjunto  $Z_u = (Z_u^1, Z_u^2, \dots, Z_u^n)$  para cada usuário.

### 2.3.2.4 Exclusão das postagens que não se configuram deslocamentos

Para se configurar um deslocamento era necessário que o usuário tivesse realizado pelo menos duas postagens ao longo de um dia ( $n \geq 2$ ) em locais diferentes, sendo excluídos da amostra todos os indivíduos que fizeram apenas uma postagem em determinado dia e aqueles que fizeram mais de uma postagem da mesma localização, ou seja, não se deslocaram.

### 2.3.2.5 Classificação das postagens

Nesta etapa realizou-se a organização e classificação (Anexo 2) das postagens a partir dos horários em que elas foram publicadas, dividindo-as em dois grupos. Postagens publicadas entre 19 h e 7 h foram consideradas de base domiciliar, partindo da premissa de que nesta faixa de horário os usuários estão em seus domicílios. Na cidade de Fortaleza um indivíduo costuma sair do seu trabalho às 18 h e leva uma média de 32 minutos do trabalho para casa (PEREIRA, 2013). Já para as postagens realizadas entre 7 h e 19 h, foi constatado que estas estavam relacionadas aos locais de atividades diárias (Trabalho, Educação e Outro) dos indivíduos.

### *2.3.2.6 Estimativa do local dos domicílios e identificação dos destinos das viagens*

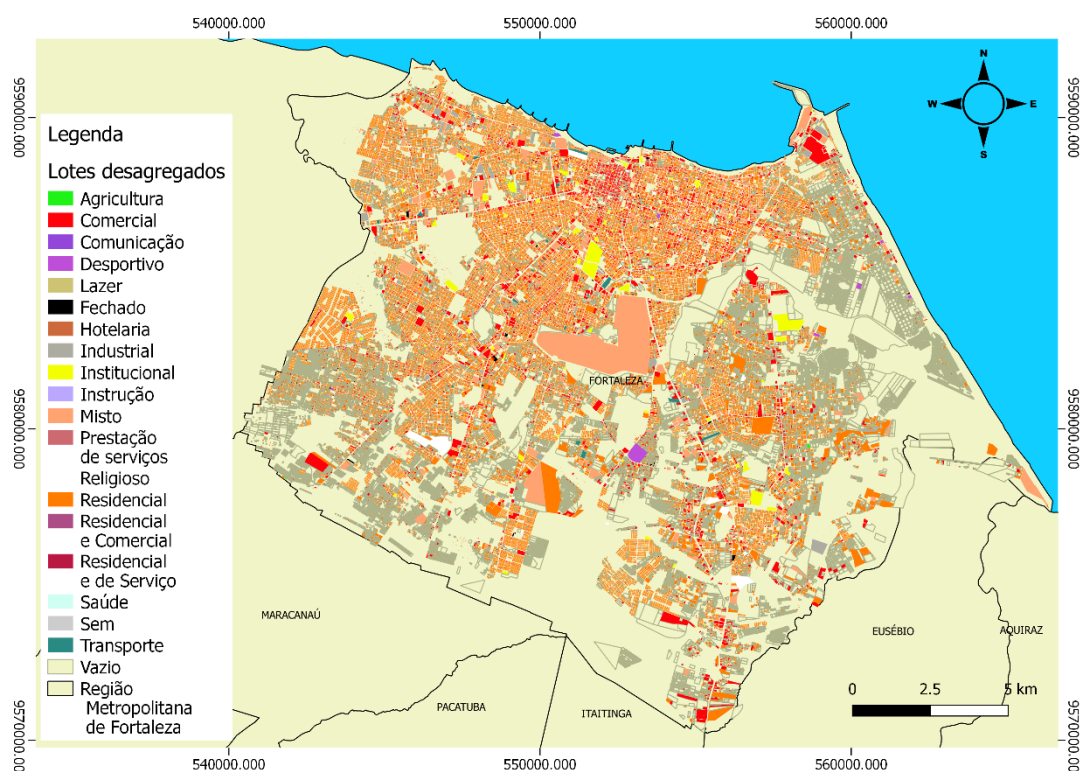
Primeiramente, realizou-se a estimativa dos locais e domicílio dos indivíduos da amostra. Para cada usuário, contabilizou-se o total de postagens entre 19h e 7h, durante oito meses, e verificou qual *ZAT* acumulou a maior quantidade delas, a qual foi considerada como a zona de domicílio do indivíduo. Já os destinos foram identificados a partir da localização das postagens realizadas entre 8h e 11h e entre 13h e 18h, intervalos nos quais verificou que esses usuários estariam realizando atividades, classificadas em Trabalho, Educação ou Outro. Além disso, desconsiderou-se o intervalo de tempo entre 12 h e 13 h, pois é nesse período que os indivíduos saem dos seus trabalhos para almoçar, dessa forma, postagens neste intervalo de tempo poderia influenciar a amostra. Finalmente, para cada indivíduo, construiu-se a conexão entre o seu domicílio e o local de cada uma das atividades identificadas, obtendo assim suas viagens. Neste caso, somente foram consideradas viagens os casos nos quais a zona do domicílio fosse diferente da zona da atividade, desconsiderado da análise possíveis deslocamentos intrazonais. Além disso, apenas levou-se em conta, como destino, a última postagem realizada pelo usuário dentro da faixa de tempo estipulada, excluindo-se as viagens em cadeia. Considerando que grande parte das viagens intrazonais são realizadas por modos não motorizados, os quais geram menor impacto na rede viária, optou-se, portanto, por desconsiderar as viagens intrazonais na análise. Ao final desse processo, foi obtida uma tabela contendo todas as viagens interzonais, com a localização das origens e destinos, realizados dentro da área de estudo.

### *2.3.2.7 Atribuição do motivo da viagem*

Construída a tabela com as informações de origem e destino de todas as viagens, passou-se à verificação da origem, se realmente é de base residencial e consequentemente à atribuição de motivos a esses deslocamentos, os quais foram classificados em Trabalho, Educação e Outro motivos. Para isso, implementou-se uma rotina na linguagem R, cujo objetivo foi classificar os pontos nas origens e destinos de acordo com o tipo de uso do solo do respectivo lote. As informações acerca do tipo de lote foram obtidas através de dados relativos ao ano de 2015, fornecidos pela SEFIN.

Nesta base, os lotes estão categorizados em: comercial, comunicação, desportivo, do lazer, fechado, hotelaria, industrial, institucional, instrução, misto, prestação, religioso, residencial, residencial e comercial, residencial e serviço, saúde, sem classificação, transporte, fechado, agricultura e trabalho e serviços. A distribuição espacial destes usos está ilustrada na Figura 6. Desconsideraram-se as viagens nas quais as origens não se encontravam em lotes do tipo residencial.

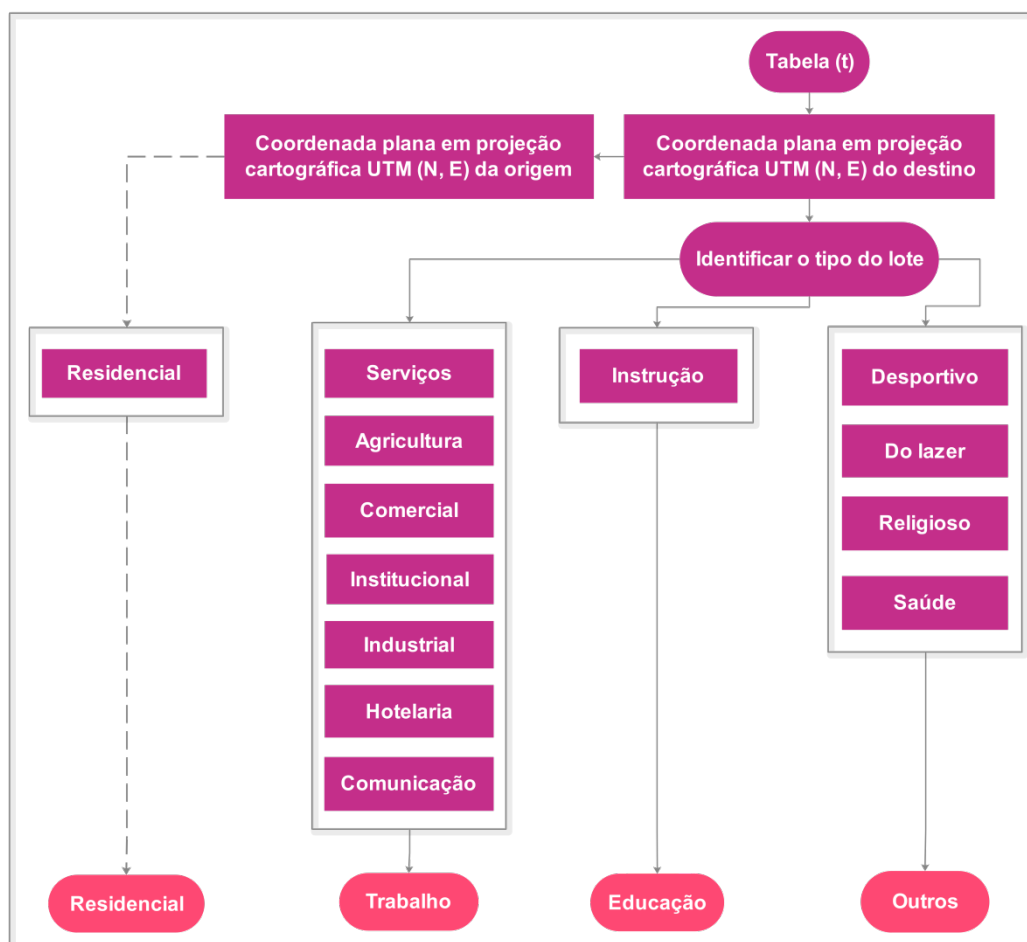
Figura 6 – Mapa de lotes desagregados do município de Fortaleza-Sistema de atividades



Fonte: Elaborado pelo autor.

Esses diversos tipos de usos do solo foram, então, agregados em quatro categorias, correspondentes às origens e aos motivos de viagem considerados nesta análise: Residencial, Trabalho, Educação e Outro. A Figura 7 detalha como foi realizada essa classificação.

Figura 7 - Algoritmo de classificação dos motivos de viagem, com base no tipo de lote da origem e do destino



Fonte: Elaborado pelo autor.

### 2.3.2.8 Obtenção das matrizes origem-destino amostrais

Com as informações de localização das origens e dos destinos das viagens, além da classificação dos motivos destas, foi possível construir três matrizes, para cada motivo de viagem, com os totais de deslocamentos  $T_{ij}$  entre as 241 zonas de análise consideradas.

### 2.3.2.9 Caracterização socioeconômica dos domicílios

Com o objetivo de se caracterizar os grupos socioeconômicos que compõem a amostra proveniente dos dados de rede sociais, foi realizada uma análise de correlação espacial entre a quantidade de domicílios que tiveram *check-ins* em uma determinada zona de análise e a quantidade de domicílios, por faixa de renda, da respectiva ZAT. As informações de renda dos domicílios foram obtidas através do trabalho de Lima (2017), que considera três estratos de renda distintos: baixa (até 3 salários mínimos), média (entre 3 e 8 salários mínimos) e alta (8 ou mais salários mínimos).

Para avaliar, espacialmente, a relação entre as duas variáveis, realizou-se, primeiramente, um teste estatístico univariado de autocorrelação espacial, com o software *GeoDA*, da variável de quantidade de domicílios, para cada estrato de renda. Com isso, foram obtidos os índices locais de associação espacial (*LISA*) para cada faixa de renda analisada, representados graficamente na forma de mapas.

Segundo Ramos (2002), a autocorrelação espacial do índice local de *Moran* é calculada com base no produto dos desvios em relação à média da variável/indicador observado. Esse tipo de gráfico de espalhamento classifica as zonas em quatro grupos, com base nos valores da zona relacionados ao valor médio de todas as zonas, cada grupo equivale ao quadrante ao qual a zona pertence no gráfico de *Moran* (LIMA, 2017). Os mapas são apresentados em quatro grupos, o formado pelas zonas *High-High*, ou seja, significa que tanto a zona como as suas vizinhas possuem um valor alto para o indicador. No caso das zonas *Low-Low*, tanto a zona como as suas vizinhas apresentam um valor baixo para o indicador. Por fim, onde essa associação é inversamente equivalente, considera-se o *High-Low*, com zonas com alto valor indicador e suas vizinhas com baixo. Finalmente, tem o último caso do grupo que são as zonas *Low-High*, ou seja, as zonas com baixo valor indicador e suas vizinhas com alto valor indicador (ANSELIN, 1995).

Após a classificação dos mapas de espalhamentos, em seguida, construiu-se um mapa de densidade *Kernel* com a distribuição espacial dos domicílios da amostra nos quais foram registrados *check-ins*. Esses mapas foram, então, comparados visualmente de modo a identificar a relação entre a concentração desses domicílios e a distribuição espacial dos distintos grupos socioeconômicos. Após essa análise visual, realizou-se ainda um teste de regressão espacial (*spatial lag*) entre as variáveis, de modo a quantificar

a correlação espacial entre a quantidade de domicílios representativos da amostra de *check-ins* e a quantidade de residências de cada faixa de renda.

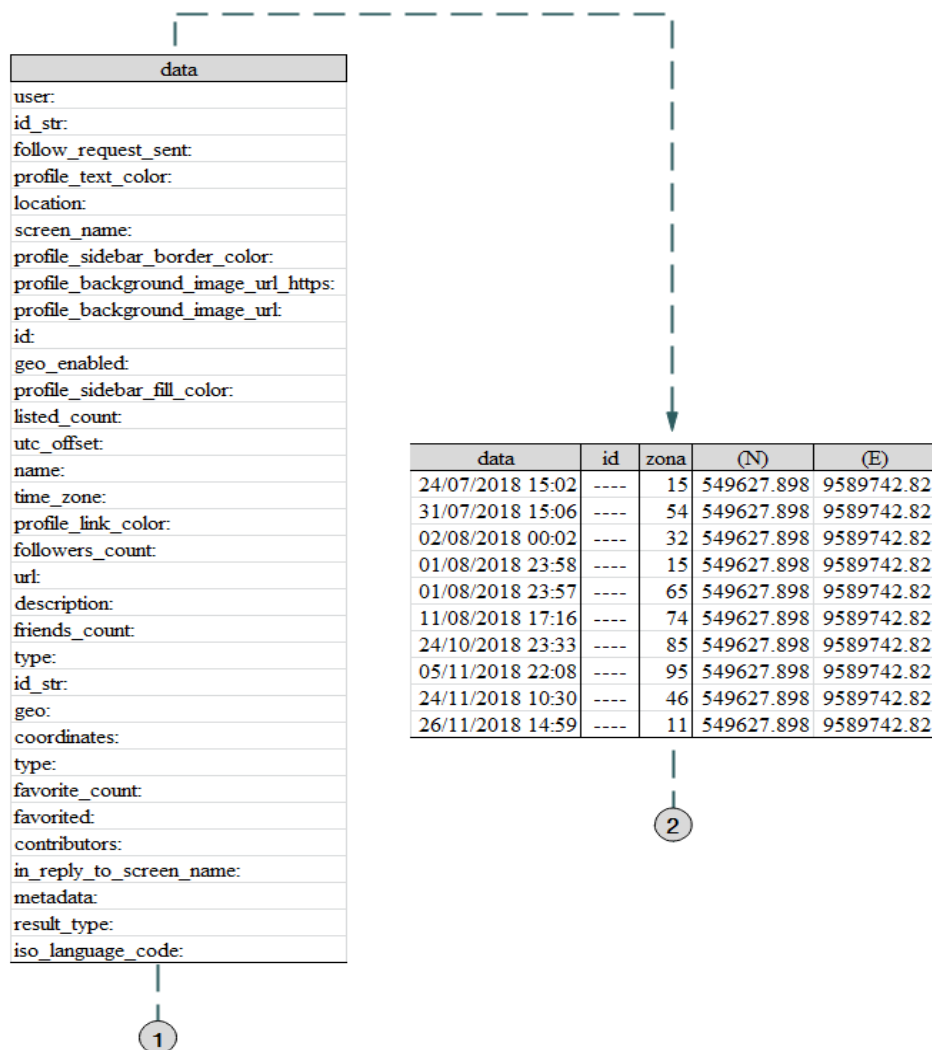
## **2.4 Resultados e discussões**

Nesta seção serão apresentados e analisados os resultados da aplicação do método proposto para o município de Fortaleza.

### **2.4.1 Obtenção e leitura dos dados**

Para a realização da análise proposta neste trabalho coletou-se uma amostra de 1,4 milhões de *check-ins*, oriundos do *Twitter* e do *Instagram*, referentes ao período compreendido entre fevereiro a junho e de agosto a outubro de 2018. Para a composição da amostra, só foram consideradas as observações relativas aos dias úteis da semana, de segunda a sexta-feira.

O conjunto de dados brutos oriundos do *Twitter* foi obtido no formato JSON (modelo de armazenamento de informações em formato de texto). Durante o período definido para a coleta, foram obtidos 30 mil *check-ins* de 10214 usuários distintos. A Figura 8 ilustra todas as informações contidas no dado bruto, muitas das quais são excluídas após uma etapa de filtragem. Com isso, permanecem apenas as informações consideradas relevantes para a análise, apresentadas no método proposto.

Figura 8 – Exemplo de informações coletadas através do *Twitter*

Fonte: Elaborado pelo autor.

Da mesma maneira, os dados do *Instagram*, coletados de forma automatizada através do site *Netlytic*, passaram pelo mesmo processo de filtragem no algoritmo, de forma a se obter apenas as informações pertinentes (Figura 9). A amostra final, coletada entre fevereiro a junho e de agosto a outubro de 2018, foi composta por 1,37 milhão de *check-ins*, associados a 131943 indivíduos. Esse total de indivíduos equivale a aproximadamente 5% da população de Fortaleza, sendo um percentual significativo quando comparado ao tamanho de uma amostra proveniente de uma pesquisa OD domiciliar.

Figura 9 - Informações coletadas através do *Instagram*

guid	latitude	longitude	link	pubdate	author	title	description	like_count	filter
---	-3.785	-38.503	---	26/01/2018 08:38	---	---	---	---	---
---	-3.785	-38.526	---	26/01/2018 08:38	---	---	---	---	---
---	-3.762	-38.48	---	26/01/2018 08:38	---	---	---	---	---
---	-3.795	-38.483	---	26/01/2018 08:38	---	---	---	---	---
---	-3.797	-38.498	---	26/01/2018 08:37	---	---	---	---	---
---	-3.756	-38.492	---	26/01/2018 08:36	---	---	---	---	---
---	-3.786	-38.567	---	26/01/2018 08:36	---	---	---	---	---
---	-3.807	-38.523	---	26/01/2018 08:35	---	---	---	---	---
---	-3.832	-38.522	---	26/01/2018 08:35	---	---	---	---	---

data	id	zona	(N)	(E)
23/07/2018 09:27	----	24	562646.618	9582168.84
07/12/2018 17:12	----	24	562626.419	9582607.21
09/06/2018 18:30	----	32	562529.497	9583392.79
09/06/2018 18:28	----	32	562529.497	9583392.79
01/07/2018 12:29	----	32	562529.497	9583392.79
31/07/2018 14:05	----	32	562442.392	9583397.43
30/09/2018 06:43	----	32	562441.612	9583397.28
13/06/2018 22:24	----	32	562413.628	9583408.76
11/08/2018 09:21	----	32	562413.628	9583408.76
12/08/2018 13:31	----	32	562413.628	9583408.76

Fonte: Elaborado pelo autor.

#### 2.4.2 Atribuição das viagens e construção das matrizes origem-destino

As duas tabelas obtidas, com as informações do *Twitter* e do *Instagram*, foram unificadas, resultando em uma única tabela, denominada aqui Tabela total. Esta tabela contém apenas os dados de indivíduos que realizaram pelo menos duas postagens ao longo de um dia, como especificado no método. Logo, cada linha corresponde a uma viagem individual, com informações do id do usuário, zona de origem e de destino e horário do *check-in* na origem e no destino. Ao todo, foram obtidas 288264 viagens correspondentes a um total de 77215 indivíduos.

Para cada uma destas viagens, foi verificado se a origem era proveniente de um lote residencial e, posteriormente, determinado um motivo, a partir da identificação do tipo de uso do solo das coordenadas geográficas das origens e dos destinos. Para cada



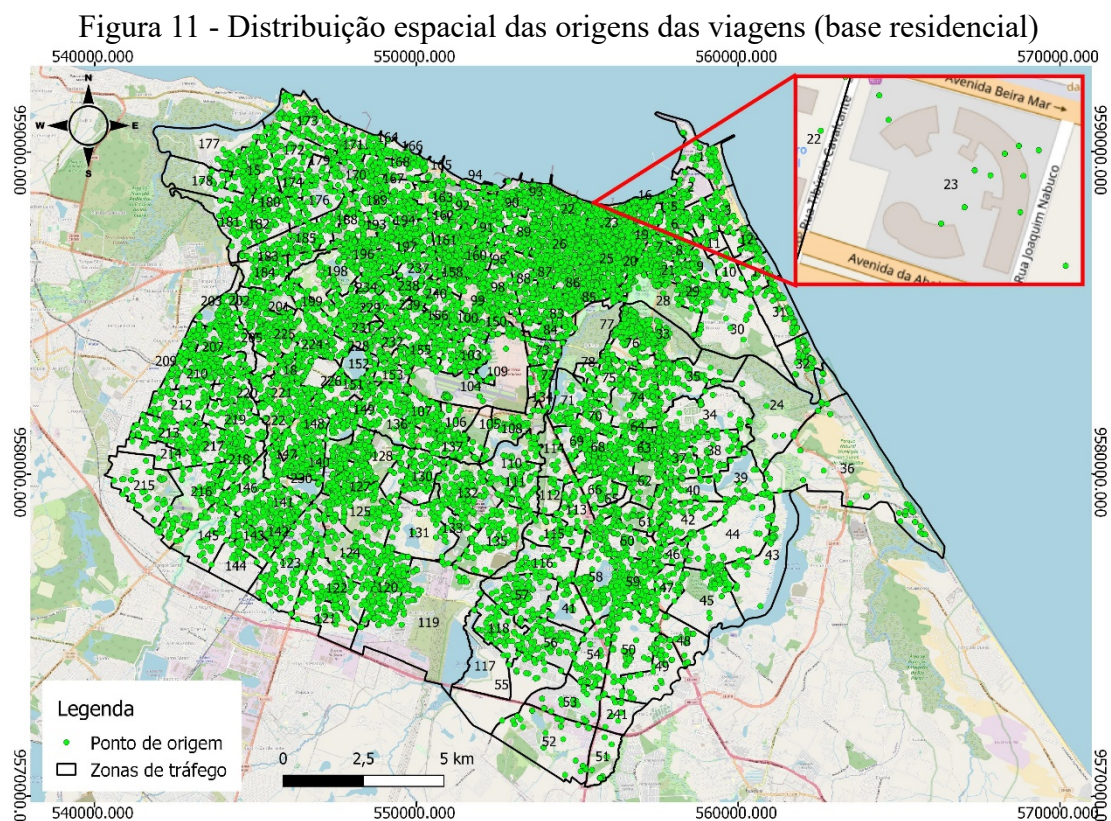
tipo de lote foi atribuído um motivo, sendo atribuído para a origem o motivo Residencial e, para o destino, Trabalho, Educação ou Outro, de acordo com a Figura 7. A Figura 10 apresenta uma parte da tabela total, com a listagem das viagens, já classificadas por motivo.

Figura 10 - Classificação a partir do lote de origem e destino

id	origem	N (O)	E (O)	lote1	destino	data	N (D)	E (D)	diasemana	lote2	motivo
---	68	555277	9580856	residencial	35	03/03/2018 01:12	557643	9582207	6	comercial	trabalho
---	188	547987	9588012	residencial	187	03/03/2018 01:10	548142	9587154	5	comercial	trabalho
---	188	547987	9588012	residencial	187	03/03/2018 01:10	548142	9587154	5	comercial	trabalho
---	31	561195	9585805	residencial	159	29/01/2018 01:18	550973	9586644	6	comercial	trabalho
---	31	561195	9585805	residencial	159	03/01/2018 03:32	550982	9586645	4	comercial	trabalho
---	93	553418	9588687	residencial	32	29/01/2018 01:45	562315	9583519	4	comercial	trabalho
---	93	553418	9588687	residencial	197	03/03/2018 01:10	549787	9587239	6	transportes	trabalho
---	93	553418	9588687	residencial	197	29/01/2018 01:18	549787	9587239	2	transportes	trabalho
---	35	557573	9583395	residencial	55	29/01/2018 01:18	554553	9573834	2	transportes	trabalho
---	35	557573	9583395	residencial	89	03/03/2018 01:10	553498	9587450	4	transportes	trabalho
---	110	552877	9579351	residencial	23	29/01/2018 01:18	556773	9587215	2	religioso	outros
---	110	552877	9579351	residencial	88	03/03/2018 01:10	553506	9585726	5	religioso	outros

Fonte: Elaborado pelo autor.

As viagens realizadas tiveram origem de base residencial. Essas estão exemplificadas na Figura 11. Percebe-se que a distribuição das origens dessas viagens se intensifica em áreas de média e alta renda, como exemplificação houve pontos localizados dentro do residencial *Landscape*. Esse residencial é considerado de luxo e está localizado na zona 23, no norte da cidade. A zona 23 e as suas vizinhas são adensadas por residências e prédios de alto padrão construtivo.

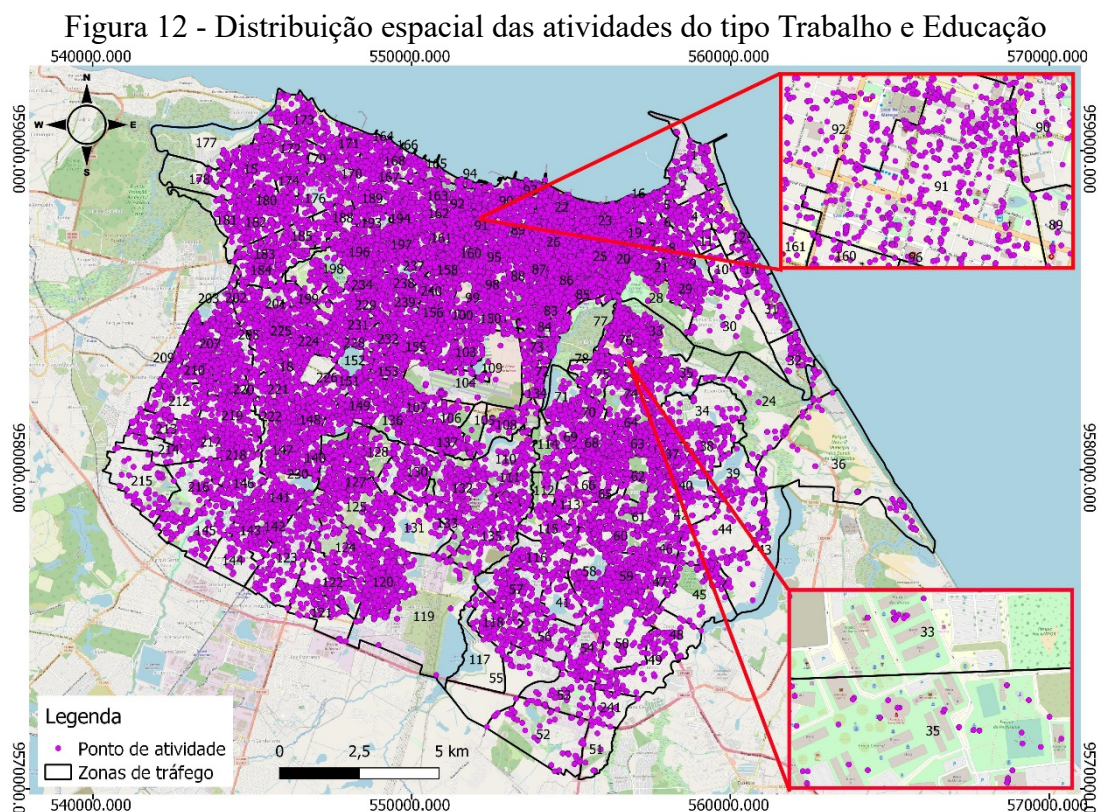


Fonte: Elaborado pelo autor.

Com as informações de viagens, diferenciadas por motivo, foram construídas três matrizes origem-destinos para a cidade de Fortaleza. Neste sentido, alguns pontos merecem destaque acerca dos resultados obtidos. Primeiramente, constatou-se que mais da metade das viagens identificadas (55%) foram por motivo Trabalho, sendo essas, realizadas a partir da premissa de que das 8 h às 11 h e das 13 h às 18h os usuários normalmente estão nos seus trabalhos. Por outro lado, as viagens por motivo educação representaram apenas 5% da amostra obtida, enquanto os 40% restantes referiram-se às viagens pelo motivo Outro. Isso indica uma tendência de que esses *check-ins* estejam representando mais fortemente viagens do tipo Trabalho, ou seja, os usuários costumam realizar publicações quando estão em seus trabalhos.

A Figura 12 ilustra a distribuição espacial das atividades compulsórias, tipo Trabalho e Educação. Verifica-se, pela análise do mapa, uma forte concentração destas atividades na região do centro comercial e na área leste da cidade. A primeira se caracteriza por ser uma região onde predominam estabelecimentos comerciais e de prestação de serviços, concentrando a maior parte dos empregos ofertados. Já na região

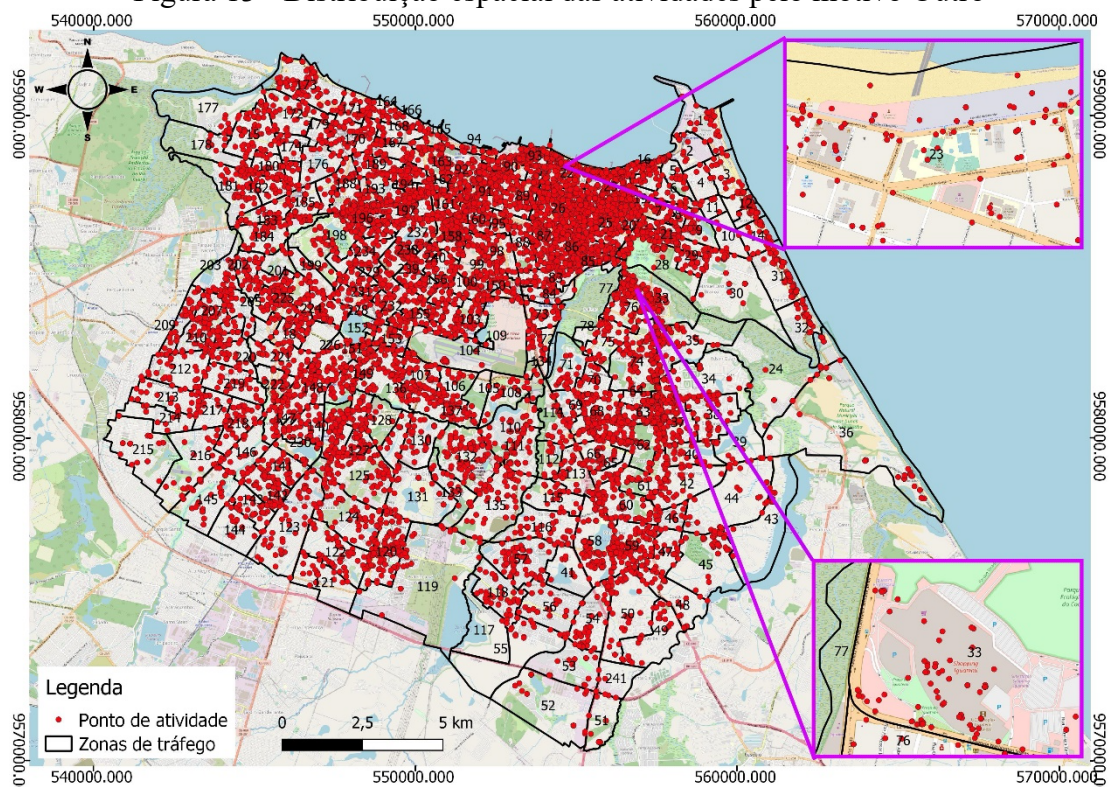
leste, ocorre uma forte concentração de faculdades particulares, o que tende a atrair muitas viagens por motivos educacionais.



Fonte: Elaborado pelo autor.

Já a distribuição espacial das atividades realizadas pelo motivo Outro pode ser observada na Figura 13. Onde percebe-se uma concentração deste tipo de atividades nas regiões mais ricas da cidade, em especial nas zonas 33 (Edson Queiroz) e 23 (Meireles). A primeira se destaca pela presença de alguns *shopping centers*, enquanto a segunda fica localizada próximo à orla de praia, atraindo uma grande quantidade de indivíduos buscando atividades de lazer.

Figura 13 - Distribuição espacial das atividades pelo motivo Outro

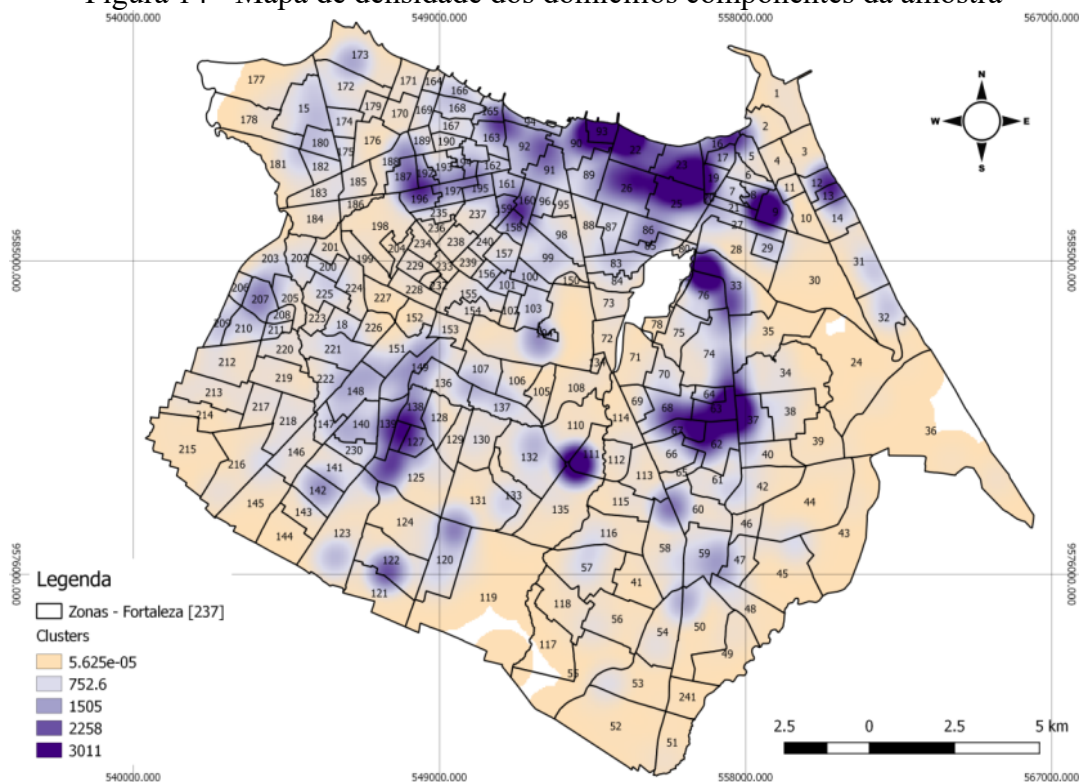


Fonte: Elaborado pelo autor.

### 2.4.3 Caracterização socioeconômica dos domicílios

Identificadas as viagens realizadas, por motivo, optou-se por fazer uma análise da situação socioeconômica dos domicílios dos indivíduos amostrados, com base na renda domiciliar. Inicialmente construiu-se um mapa de densidades da distribuição espacial dos domicílios nos quais foram realizados *check-ins* durante o período de coleta. Essa caracterização está ilustrada no mapa da Figura 14. Visualmente é percebido que há uma concentração destes domicílios na região do centro da cidade, ao norte, com alguns aglomerados mais o sul, sudeste e sudoeste.

Figura 14 - Mapa de densidade dos domicílios componentes da amostra



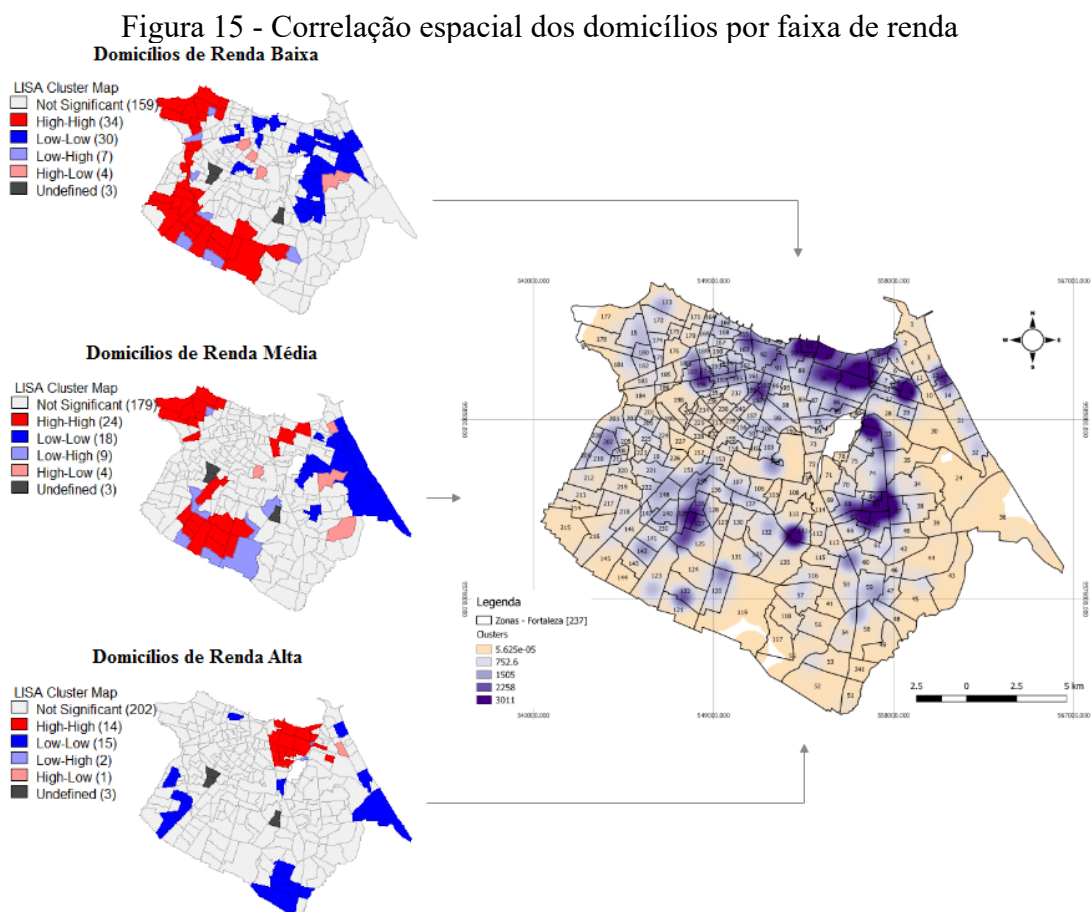
Fonte: Elaborado pelo autor.

Analisando a autocorrelação espacial da distribuição dos domicílios Figura 15, concluiu-se que as maiores densidades de domicílios da amostra de *check-ins* coincidem, de forma mais significativa, com os aglomerados de domicílios de renda média e alta. Então, as variáveis explicativas como tipos de domicílios por faixa de renda foram exemplificadas através do mapa de espalhamento de *Moran*.

Dessa forma, as zonas *High-High* que possuem em sua maioria domicílios de renda baixa estão clusterizadas ao sudoeste, oeste e noroeste. Por outro lado, as zonas *High-High* de renda média clusterizam a noroeste, sudoeste e nordeste, e por fim, as zonas que possuem em sua maioria domicílios de renda alta estão clusterizadas a norte e nordeste exclusivamente. Em contrapartida, ao analisar a variável renda baixa, foi observado que as zonas *Low-Low* estão a noroeste e nordeste, enquanto os domicílios de renda média que pertencem ao grupo *Low-Low* se encontram a leste exclusivamente. No entanto, as zonas *Low-Low* de renda alta se localizam a oeste, sul e leste.

A partir dessa análise visual entre o mapa de espalhamento de *Moran* (*Moran Map*) e o mapa da densidade dos domicílios, percebe-se que existe uma forte relação entre as zonas constituídas por domicílios de média e alta renda com os *clusters* de domicílios

“*check-ins*” da amostra. A partir dessa constatação, cabe assumir, portanto, que a amostra obtida, proveniente de dados de redes sociais, representa, de forma pouco significativa, as viagens dos grupos pertencentes a estratos da população com renda menores.



Fonte: Elaborado pelo autor.

Para avaliar de forma quantitativa a relação entre os domicílios da amostra e a distribuição das residências por faixa de renda, realizou-se uma regressão espacial do tipo *spatial lag*. Pelos valores do coeficiente de correlação espacial  $R^2$ , observa-se, de fato, uma maior correlação espacial entre a localização dos domicílios da amostra com a localização dos domicílios de média e alta renda, com valores de  $R^2$  de 0,4 e 0,5 respectivamente. Enquanto os domicílios de baixa renda apresentaram um coeficiente de correlação espacial de 0,28.

Uma das hipóteses levantadas para explicar a menor representatividade dos domicílios de baixa renda na amostra de viagens, proveniente de dados de redes sociais, é o fato da dificuldade de acesso à internet por parte dos indivíduos deste grupo, apesar

de que nos últimos anos houve uma maior disseminação dos celulares para a maioria da população de baixa renda. Segundo dados do IBGE de 2018, apenas 65,9% dos domicílios com renda menor ou igual a três salários mínimos têm internet em casa, seja móvel ou fixa.

## 2.5 Considerações Finais

Neste trabalho, foram identificados padrões de deslocamento em Fortaleza através de dados obtidos de redes sociais (*Twitter e Instagram*). A partir dos *check-ins* georreferenciados, realizados pelos usuários desses aplicativos, foi possível identificar suas viagens cuja base fosse o domicílio. Além disso, com o auxílio da base de dados de uso do solo desagregada ao nível do lote, foi possível ainda identificar os motivos dessas viagens, discriminados em Trabalho, Educação ou Outro. Por fim, realizou-se uma análise de caracterização socioeconômica, com base na renda, dos domicílios que compuseram a amostra. Diante dos resultados obtidos, foi possível concluir que a base de dados provenientes de redes sociais georreferenciadas se apresenta como uma alternativa viável aos métodos tradicionais de obtenção de dados de viagens, em especial as pesquisas OD domiciliares, de alto custo de obtenção e estacionários no tempo.

Apesar das vantagens identificadas na utilização deste tipo de dado, foram encontradas também algumas limitações. Uma delas foi a pequena representatividade de domicílios de faixa de renda mais baixa na composição da amostra de indivíduos. Com isso, os padrões de deslocamento observados só refletem as características de uma parcela com níveis de renda mais elevados. Além disso, verificou-se uma elevada proporção de viagens do tipo Outro e poucas viagens por motivos educacionais. Isso leva a crer que os usuários de redes sociais tendem, de modo geral, a realizar *check-ins* quando estão realizando atividades que não sejam compulsórias.

No entanto, o trabalho proposto possui uma limitação no que se refere à representatividade dos dados (comportamento da população como um todo). Pois, foi observada que boa parte da população ainda não tem acesso às redes sociais baseadas em localizações, ou seja, indivíduos de baixa renda. Essa limitação, no entanto, exigirá novas pesquisas. Para estudos futuros, propõe-se um estudo sobre a variabilidade de viagens, além das etapas de geração e distribuição, e dessa forma, conseqüentemente, uma comparação entre a matriz OD proposta pelo método e uma matriz externa (gerada por

alguma pesquisa). Portanto, os resultados amostrais obtidos só são expansíveis em sua maioria para viagens não compulsórias de determinados grupos socioeconômicos, de faixa de renda mais elevada.



### 3 ANÁLISE DE VARIABILIDADE DOS DESLOCAMENTOS URBANOS A PARTIR DE DADOS DE REDES SOCIAIS EM FORTALEZA

#### 3.1 Introdução

De acordo com Zhong *et al.* (2015) existe uma variabilidade, entre os dias da semana, entre indivíduos e entre lugares, nos padrões de mobilidade em uma rede de transportes. Esse fenômeno ocorre por que as necessidades dos usuários não são as mesmas de um dia para o outro (SCHLICH e AXHAUSEN, 2003). A compreensão desse comportamento permite a obtenção de uma visão mais precisa e direta dos padrões espaço-temporais, além de possibilitar o estudo do comportamento de viagens diariamente em relação aos diferentes motivos de viagens. Certos estudos como o de Zhong *et al.* (2015) consideram a variabilidade nos padrões de mobilidade, como um fator importante para identificar a diversidade no comportamento de viagens das pessoas no meio urbano.

De acordo com Schlich e Axhausen (2003), uma das principais limitações para efetuar análise de variabilidade de dados relativos a viagens é a indisponibilidade de dados amostrais grandes o suficiente para tal análise. No entanto, ultimamente, a facilidade e o baixo custo de coletar grandes volumes de dados sobre mobilidade urbana em uma escala razoável de tempo contribuiu com o crescente aumento no interesse científico em pesquisas relacionadas ao tratamento de dados de mobilidade (ZHONG *et al.*, 2015).

A importância de compreender esse comportamento é reconhecida por e Jones e Clarke (1988) e Bartosiewicz e Pielesiak (2019) que ressaltam que o conhecimento da variabilidade de viagens é fundamental para o sucesso no processo de tomada de decisão do planejamento urbano de transportes, pois a análise da variabilidade diária do comportamento de viagens permite que pesquisadores sociais e modeladores da demanda por viagens alcancem melhores resultados com a utilização de ferramentas estatísticas para compreender melhor o comportamento de viagens dos indivíduos.

Dessa forma, este capítulo tem como objetivo investigar o grau de variação do comportamento de viagens dos usuários do sistema de transportes em Fortaleza utilizando dados de redes sociais. Para isso, este capítulo está organizado em cinco Seções, sendo esta introdução a primeira delas. Na Seção 3.2, na revisão da literatura

analisa-se trabalhos anteriores sobre a variabilidade a fim de encontrar as vantagens e limitações dos métodos adotados, tanto para coleta de dados como para o tratamento. Além disso, nessa sessão serão também introduzidos conceitos fundamentais sobre a variabilidade comportamental dos usuários no que se diz respeito às viagens que eles realizam. Na Seção 3.3, é apresentado o método adotado para efetuar a análise de variabilidade. Em seguida, na Seção 3.4, o método é aplicado e são apresentados os resultados da análise, e, por último, na Seção 3.5 discute-se sobre as contribuições metodológicas e fenomenológicas deste estudo.

## **3.2 Revisão bibliográfica**

### ***3.2.1 Variabilidade comportamental dos usuários***

A principal causa do problema de medir a variabilidade tem sido a dificuldade na obtenção dos dados sobre o comportamento de viagens dos entrevistados (SCHLICH e AXHAUSEN, 2003). Atualmente há disponibilidade de grande volume de dados, principalmente advindos das redes sociais. Isso permitiu que vários trabalhos realizassem a análise de mobilidade dos indivíduos em diversas cidades do mundo em uma escala sem precedentes.

Koppelman e Pas (1984) classificam a variabilidade do comportamento de viagens ao longo de uma sequência de dias em: padrões intrapessoal e interpessoal. A variabilidade intrapessoal diz respeito à variação no comportamento cotidiano de uma determinada pessoa, sendo esse sistemático, onde as diferenças são ocasionadas pelo dia da semana. Por outro lado, a variabilidade interpessoal refere-se a diferenças no comportamento de indivíduos (no mesmo dia ou em dias diferentes) (PAS e SUNDAR, 1995).

Para essas análises de variabilidade é necessário ainda obter dados referentes ao comportamento de viagens dos usuários. Esses dados são classificados como transversais ou longitudinais. A análise dos dados de viagem transversal considera a viagem de um determinado indivíduo como habitual (Hanson e Huff, 1981), enquanto dos dados longitudinais são usados para detectar mudanças comportamentais dos usuários ao longo do tempo (ELANGO *et al.*, 2007).

Os dados de viagem transversal podem apresentar problemas na obtenção de inferência a partir dos dados, devido à variabilidade temporal no comportamento de viagens. Dados de viagens transversais têm limitações analíticas em relação à dinâmica comportamental da viagem, onde não é possível capturar a persistência de hábito (por exemplo, indivíduo persiste em realizar compras em um local mesmo depois de abrir um local mais próximo com a mesma utilidade) (PENDYALA e PAS, 2000).

Já os dados longitudinais, por outro lado, são importantes em termos de política e análise (ELANGO *et al.*, 2007). Eles facilitam a identificação de relações de causa e efeito, pois podem explicar a dinâmica comportamental da viagem (PENDYALA e PAS, 2000). Além disso, são eficientes para coleta em caso de pesquisas em painel e podem fornecer benefícios em relação ao custo da pesquisa e a eficiência dos parâmetros. No entanto, o processo de coleta de dados longitudinais tem suas limitações. Quando coletados passivamente exigem tecnologia de ponta e mão de obra qualificada, além disso, depende de tecnologias externas, tais como: wireless e GPS (Sistema de Posicionamento Global) (ELANGO *et al.*, 2007).

As diferentes formas de dados têm como propósito compreender o comportamento das viagens rotineiras. A questão de como as viagens se repetem ao longo de dias tem sido objeto de investigação há anos (SCHLICH e AXHAUSEN, 2003). Há na literatura, uma série de publicações acerca do estudo da variabilidade de viagens, sendo que, esses trabalhos foram desenvolvidos a partir de dados oriundos de métodos tradicionais e alternativos. Os dados provenientes de técnicas de coleta por métodos tradicionais, em sua maioria são compostos por informações detalhadas da viagem, diferente dos métodos alternativos, onde são coletados principalmente de forma passiva através de tecnologias como GPS, Bluetooth e chamadas telefônicas por meio de redes GSM (Sistema Global para Comunicações Móveis).

### ***3.2.2 Variabilidade de viagens através de métodos tradicionais***

Os estudos a respeito da variabilidade de viagens tiveram início na década de 60, do século XX, poucos trabalhos anteriormente tentaram estudar o assunto. Marble e Bowlby (1968) examinou o nível de repetição das viagens para diferentes propósitos através da aplicação de diários de viagem. O estudo foi realizado na cidade de Cedar Rapids, no estado de Iowa nos Estados Unidos. Os autores coletaram dados por aproximadamente um mês e descobriu que, para a maioria das famílias observadas, cerca

de 75% delas possuíam padrões e repetições de viagens. Porém, não foi estudada a repetição ou a variabilidade em um nível individual.

Pas (1986) investigou o fenômeno acerca da variabilidade intrapessoal na cidade de Reading, na Inglaterra. Para isso, ele utilizou os dados que foram coletados em outra pesquisa referente a janeiro e março de 1973, durante a *Activity Diary Survey* (SHAPCOTT, 1978). Os dados foram coletados através de uma entrevista, através da qual solicitou-se a cada um dos entrevistados registrasse as origens e destinos de seus deslocamentos diários durante um período de cinco dias. A amostra de entrevistados consistiu-se de 136 pessoas compondo os dados do *Reading Activity Diary Survey*. Cada pessoa forneceu o número de viagens realizadas em cada um dos cinco dias da semana consecutivos, bem como o gênero, renda da família e se possuía emprego. Os resultados mostraram que o uso de uma amostra de vários dias depende da estrutura de custos da coleta de dados e da proporção da variabilidade total atribuída as diferenças entre as pessoas, além disso, o número ideal de dias consecutivos pode ser de aproximadamente dois dias para um levantamento a respeito do comportamento de viagens a ser utilizado na estimativa de modelos de geração de viagem diária.

Pas e Sundar (1995) estudaram a variabilidade intrapessoal diária em viagens urbanas, utilizando um conjunto de dados coletados na cidade de Seattle, no estado de Washington nos Estados Unidos. Os dados foram provenientes de uma pesquisa OD domiciliar, entrevistando um total de 186 famílias. Os resultados mostraram que a proporção da variabilidade total em relação ao comportamento individual é atribuível à variabilidade intrapessoal. Nos domicílios de duas pessoas, a variabilidade intrapessoal foi de aproximadamente 45% da variabilidade total na frequência de diária de viagens. Além disso, a variabilidade intrapessoal foi responsável por mais de 70% da variabilidade dentro do agregado familiar composto por duas pessoas. De uma forma geral, com relação a certos aspectos de viagem diária, há uma considerável variação no comportamento de viagens dos indivíduos.

Susilo e Kitamura (2005) estudaram a variabilidade do dia a dia dos residentes das cidades de Karlsruhe e Halle, Alemanha. Esse estudo se deu através de um conjunto de dados conhecidos como *Mobidrive* (conjunto de dados longitudinais que serviu de base para vários trabalhos a respeito de variabilidade), que foram obtidos em 1999. Essa amostra foi referente a 317 indivíduos com idade superior a seis anos, compondo um total de 139 domicílios. A análise focou na associação entre a extensão e a variabilidade do espaço de ação de um indivíduo. Esse estudo confirmou, como já era esperado, que

indivíduos que trabalham fora de casa ou que estudam, tendem a realizar as mesmas viagens diariamente, ou seja, nos dias da semana quando as atividades tendem a serem obrigatórias e rotineiras, além disso, as atividades tendem a ser fixas e o espaço de ação tende a ser recorrente.

Tarigan e Kitamura (2009) quantificaram a frequência de viagens do tipo lazer e a variabilidade dessas viagens ao longo de semanas. A frequência e a variabilidade das viagens foram medidas para seis tipos de atividades: contato social, esportes, recreação, cultura, natureza e compras. Os dados utilizados nessa pesquisa referem-se ao *Mobidrive*, uma pesquisa que ocorreu nas cidades de Halle e Karlsruhe, na Alemanha como apresentado em (SUSILO e KITAMURA, 2005). Os resultados corroboram a hipótese de que os desejos e as necessidades dos indivíduos são mais variáveis e complexos se estiverem envolvidos nas atividades do tipo esporte, contato social e recreação. Os autores ainda observaram que não há correlação estatística entre o tempo de viagem e a variabilidade em certas atividades, como as do tipo cultura, natureza e compras. Dessa forma, os resultados sugerem que ambas as variáveis são independentes e que indivíduos resolvem realizar atividades com base em experiências anteriores bem sucedidas de viagens a lazer.

Raux e Cornelis (2016) identificaram como as pessoas organizam as suas atividades diárias e como ocorrem as variações temporais com relação aos padrões de atividade individuais. Essa variabilidade foi analisada de acordo com vários períodos de tempo dentro de uma semana, identificando as variações interpessoais e intrapessoais de acordo com os atributos oscilantes das viagens por diferentes tipos de atividades. Os autores utilizaram para esse estudo dados referentes a uma pesquisa de mobilidade baseada em um método tradicional de entrevista a domicílio. Foram entrevistados 717 indivíduos durante um espaço de tempo de sete dias, o que resultou em 4000 pessoas entrevistadas. Essa pesquisa ocorreu na cidade de Ghent (Bélgica) em 2008, onde a taxa de resposta global foi de 18%. Diferentes métodos de indicadores de variância e de alinhamento sequenciais foram aplicados para a medição da variabilidade do comportamento dos deslocamentos. Os autores perceberam que há uma maior diferença no número de viagens por dia ao longo da semana para um determinado indivíduo do que para outros, e que a variabilidade intrapessoal é alta durante a semana quando comparada exclusivamente com o sábado e o domingo. Por outro lado, a variabilidade total na alocação de tempo é aproximadamente constante, quaisquer que sejam os períodos considerados na semana, exceto nos fins de semana onde há uma queda.

Os esforços de pesquisa para compreender as variações cotidianas no comportamento de viagens, variação intrapessoal e interpessoal, podem ser vistos nos trabalhos citados anteriormente. Esses estudos mostraram que há uma variabilidade significativa no comportamento das viagens para diferentes tipos de atividades como os trabalhos de (SUSILO e KITAMURA, 2005; TARIGAN e KITAMURA, 2009).

Estudos relacionados à variabilidade de viagens, como Pas (1986), dizem que o comportamento de viagem pode variar de um dia para o outro. Outros expõem a diferença entre dias da semana e fins de semana, como (RAUX e CORNELIS, 2016). Pas (1986) e Raux e Cornelis (2016) buscaram compreender a variabilidade entre indivíduos (interpessoal) e por indivíduos separadamente. Os trabalhos vistos na literatura apresentam diferentes conclusões. Os métodos de coletas de dados vistos nessas pesquisas são advindos de métodos tradicionais, como: pesquisa OD domiciliar e diários de viagem. E esses, de certa forma podem dificultar e influenciar os resultados.

As pesquisas tradicionais, como exemplo a pesquisa OD domiciliar, coletam um ou dois dias de informações sobre as viagens realizadas por cada domicílio. Essas pesquisas tradicionais fazem um trabalho de capturar e compreender o comportamento médio das viagens de cada morador do domicílio, além de colher informações sociodemográficas. Entretanto, essas pesquisas não fornecem dados suficientes para análises comportamentais mais detalhadas em nível desagregado (PENDYALA e PAS, 2000). Além disso, como se pode observar através dos métodos de coleta de dados dos trabalhos apresentados, essas pesquisas podem se tornar muito caras e podem demandar muito tempo para que os dados estejam prontos para efetuar a análise comportamental dos indivíduos.

### ***3.2.3 Variabilidade de viagens através de métodos alternativos***

Com a evolução da análise comportamental de viagens e a otimização de técnicas estatísticas, ocorre uma maior necessidade de avaliar a variabilidade diária das viagens em níveis intrapessoal e interpessoal. A variabilidade diária de viagens é resultado do processo natural de cada indivíduo de se deslocar a partir de um determinado modo de transporte para realizar alguma atividade, o que acaba de certa forma afetando o sistema. Os métodos alternativos de obtenção de dados em contrapartida aos métodos tradicionais podem fornecer informações detalhadas e de baixo custo a respeito da

variabilidade de viagens dos habitantes de uma região. Diversos estudos utilizaram esses dados para estudar e compreender a variabilidade de viagens.

Stopher e Zhang (2011) analisaram a repetitividade e a variabilidade de viagens. Os dados utilizados nesta pesquisa são provenientes do sistema GPS, do inglês, *Global Positioning System*, os quais foram coletados no subúrbio ocidental de Adelaide, no sul da Austrália. Para esse estudo, foram escolhidos aleatoriamente 50 domicílios da mesma área geográfica, em que cada morador levou consigo um receptor GPS por 28 dias. Para obter o número de viagens foi necessário contar o número de ocasiões em que cada viagem ocorreu durante vários dias. A partir dessa contagem, os atributos de viagem foram comparados. Se as viagens seguissem um mesmo padrão de tempo, atividade e hora do dia para o começo e o final, esses atributos sugeririam que os padrões de viagens eram repetitivos e não variáveis. No entanto, se algum desses atributos fosse diferente, concluiria que os padrões de viagem eram diversos. Os autores concluíram que não há padrões repetitivos de viagens, e que a variabilidade de viagens por motivo trabalho e estudo é baixa ao longo dos dias da semana.

Motahari *et al.* (2012) investigaram como os fatores temporais influenciam nas características de viagens e na distribuição espacial. O estudo foi realizado na cidade de São Francisco, Califórnia. Para a realização desse estudo, foram coletados registros de chamadas telefônicas de aproximadamente três mil usuários, por uma média de quatro dias, sendo dois dias úteis e dois finais de semanas. A metodologia de análise foi dividida em duas etapas. Na primeira etapa foram investigados como ocorre a variabilidade de viagens dos usuários ao longo da semana e das horas do dia. Na segunda etapa, os usuários foram monitorados por um mês e após isso, feita uma análise estatística sobre a distribuição dos locais mais visitados pelos usuários. Os resultados mostraram que a escala de variabilidade é menor durante o horário de trabalho e mais dispersa nos finais de semana, além disso, testes estatísticos mostraram uma diferença significativa nos padrões de viagens.

Zhong *et al.* (2015) apresentaram um conjunto de medidas de variabilidade em níveis intrapessoal e agregados utilizando dados de cartão inteligente provenientes de viagens do transporte público da cidade de Cingapura. O sistema gerou uma média de 5 milhões de registros de viagens diários que foram coletados em abril de 2014. Os estudos para a compreensão da variabilidade foram realizados por meio de análise estatística, matriz de correlação e métodos de agrupamento baseado em rede. Após as análises, foi observado que a variabilidade dos padrões de mobilidade existe na escala individual e em

nível agregado, porém, na estrutura geral dos movimentos urbanos permanece quase a mesma ao longo da semana. Além disso, o número de viagens e os padrões de mobilidade variam de um dia para o outro.

Xianyu *et al.* (2017) realizaram um estudo a respeito da variabilidade de viagens o usando coordenadas obtidas pelo sistema GPS, disponível em smartphones. A pesquisa foi realizada em Xangai na China, onde houve uma participação de 46 pessoas. A metodologia aplicada foi dividida em duas etapas, na primeira, uma sequência multidimensional foi aplicada para medir o grau de dissimilaridade das viagens individuais entre diferentes pares. Na segunda, um modelo de regressão foi aplicado para estimar os efeitos das variáveis sociodemográficas e os dias da semana. Os resultados indicaram que em média, cada participante teve em média 1,4 viagens de atividades fora de casa, 2,6 viagens por dia da semana e 2,3 viagens no final de semana. Além disso, percebeu-se uma significativa variabilidade de viagens entre os dias da semana, sendo maior na quinta-feira e com atividades repetitivas ao longo da semana. Ademais, os homens em geral têm uma menor variabilidade intrapessoal em suas atividades diárias, isso pode ser explicado devido ao fato de que são menos propensos a serem afetados por cuidados infantis e compras. Outro fator importante é que a idade tem um efeito negativo na variabilidade intrapessoal, sendo os jovens, os mais propensos a terem uma maior variabilidade em seus padrões diários de viagens.

Os trabalhos supracitados tiveram um papel importante na compreensão da variabilidade de viagens através de dados provenientes de métodos alternativos. A partir de tecnologias como GPS, smartphones e bilhetagem eletrônica, tornou-se possível a coleta de dados advindos de vários dias, e dessa forma através de métodos estatísticos, foi possível uma maior compreensão espacial da variação de viagens tempo. No entanto, há pouco esforço de pesquisa sobre esse fenômeno no Brasil, utilizando métodos semelhantes, como os dados de viagens das redes sociais. Se deve ponderar ainda que no caso dos dados de redes sociais, esses podem não ser representativos do comportamento devido a uma questão socioeconômica muito relevante na realidade brasileira: o fato de que a internet não é de fácil acesso a todos os brasileiros. Segundo o IBGE (2018), cerca de aproximadamente 45,9 milhões de brasileiros ainda não tinham acesso à internet em 2018. Ainda com base na pesquisa, boa parte dos entrevistados declarou não saber utilizar a internet, enquanto a outra parcela declarou não ter acesso a internet por questões financeiras.



### 3.2.4 Considerações

Diante dessas informações, pode-se afirmar que para essa pesquisa, o tipo de variabilidade do comportamento de viagens dos indivíduos é interpessoal. Além disso, descrições sobre o comportamento usual dos indivíduos na cidade motivou a proposição de hipóteses para embasar o estudo em Fortaleza. Essas hipóteses se subdividem em dois grupos: hipóteses sobre o comportamento e hipóteses sobre a variabilidade desse comportamento.

As hipóteses comportamentais foram as seguintes:

- Há diferença no comportamento de viagens entre os dias da semana;
- Nas sextas-feiras, o comportamento de viagens é parecido com o dos finais de semana;
- Nos finais de semana, realiza-se mais viagens por motivo Outro do que na semana;

Enquanto as hipóteses sobre a variabilidade desse comportamento foram:

- A variabilidade de viagens por motivo Trabalho é menor no final de semana do que em dias úteis da semana;
- A variabilidade de viagens pelo motivo Outro é maior nos finais de semana do que em dias da semana;
- A variabilidade de viagens é maior em viagens pelo motivo Outro do que em viagens por motivo Trabalho;
- A variabilidade de viagens é maior na sexta-feira comparada com os demais dias úteis da semana.

### 3.3 Método

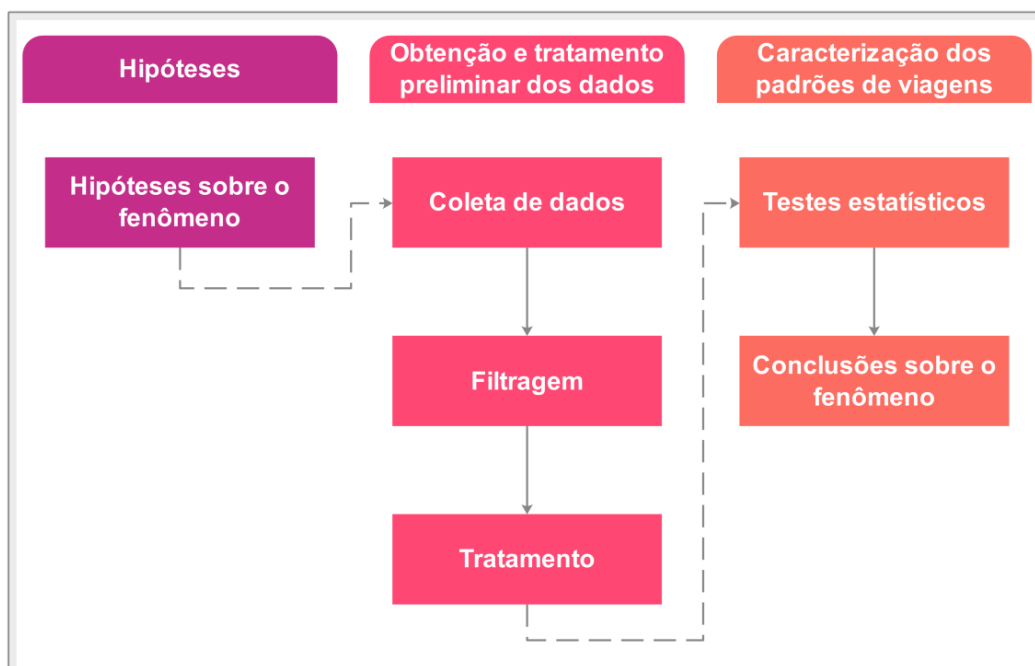
O método empregado nessa pesquisa consiste na investigação do volume de viagens entre dias (segundas a domingos) e na caracterização de padrões de deslocamento dos usuários da cidade de Fortaleza baseando-se nas hipóteses propostas.

Os dados de viagens utilizados nessa investigação foram obtidos através de informações de *check-ins* de perfis de redes sociais, tais como *Instagram* e *Twitter*. Além disso, as viagens foram classificadas apenas em dois motivos, Trabalho e Outro. Esses

motivos foram atribuídos a depender do horário e da localização onde foram realizados os *check-ins*. Esses dados foram obtidos como descrito no Capítulo 2.

Nesse trabalho, serão empregadas técnicas estatísticas, como ANOVA e Tukey, a fim de identificar e analisar a variabilidade do comportamento de viagens entre os diferentes meses, considerando apenas os dados dos dias da semana. O fluxograma da proposta metodológica deste trabalho está representado na Figura 16.

Figura 16 - Dias da semana em função de viagens ao longo dos meses de análise



Fonte: Elaborado pelo autor.

A análise de variabilidade foi dividida em três etapas: hipóteses sobre o fenômeno, obtenção e tratamento preliminar de dados e caracterização dos padrões de viagens dos indivíduos. A primeira consistirá em propor hipóteses sobre o fenômeno analisado, resultado da análise da revisão bibliográfica feita da seção anterior. Na segunda etapa, foram realizados a obtenção e o tratamento preliminar das viagens, e conseqüentemente será feita uma filtragem dos dados para capturar o comportamento típico dos usuários a fim de que seja possível investigar as hipóteses aventadas. Na última etapa ocorrerá a caracterização dos padrões de viagens dos indivíduos, e por fim as hipóteses propostas serão investigadas através da realização de testes estatísticos como ANOVA e Tukey. Como resultado dessa etapa, esperam-se conclusões importantes sobre o fenômeno em análise.

### **3.3.1 Tratamento preliminar dos dados**

De posse das hipóteses propostas e antes de efetuar as análises estatísticas, foi realizada uma filtragem preliminar dos dados do estudo da variabilidade de viagens, visando dar maior confiabilidade ao estudo. Nessa investigação, busca-se utilizar viagens de dias habituais aos usuários a fim de que se possa representar o comportamento usual desses. Para isso, serão extraídos do conjunto de dados, os dias atípicos em relação aos deslocamentos urbanos, o que retratará o comportamento típico em dias úteis e livre de influências, são eles:

- a) Dias com imprecisões (causados por algum problema, ruído na coleta de dados);
- b) Dias de feriados nacionais e regionais;
- c) Dias que antecedem festividades (carnaval, véspera de natal e réveillon).

Segundo Wright *et al.* (1997), embora o efeito dos feriados seja pequeno nas estimativas médias do volume de tráfego, não é insignificante nos coeficientes de variação. Além disso, segundo Oliveira (2004) pode-se excluir os dias nos quais houve ocorrência de desvios resultante de uma obra de infraestrutura. Assim, ao definir e aplicar esses critérios ao conjunto de dados espera-se que as medidas adotadas tragam uma maior confiabilidade nas análises quanto uma maior representatividade do comportamento típico do tráfego. Além disso, ainda com o intuito de capturar o comportamento típico das pessoas para esse trabalho, serão utilizados os dados provenientes dos meses de fevereiro a junho e de agosto a outubro. Essa estratégia é uma recomendação de Oliveira (2004), que com o mesmo propósito verificou que nesse período do ano, as pessoas exercem um comportamento semelhante e habitual. Ao finalizar essa etapa de filtragem dos dados, esses estarão prontos para serem utilizados nas análises das etapas seguintes, de investigação das hipóteses de estudo.

### **3.3.2 Caracterização dos padrões de viagens**

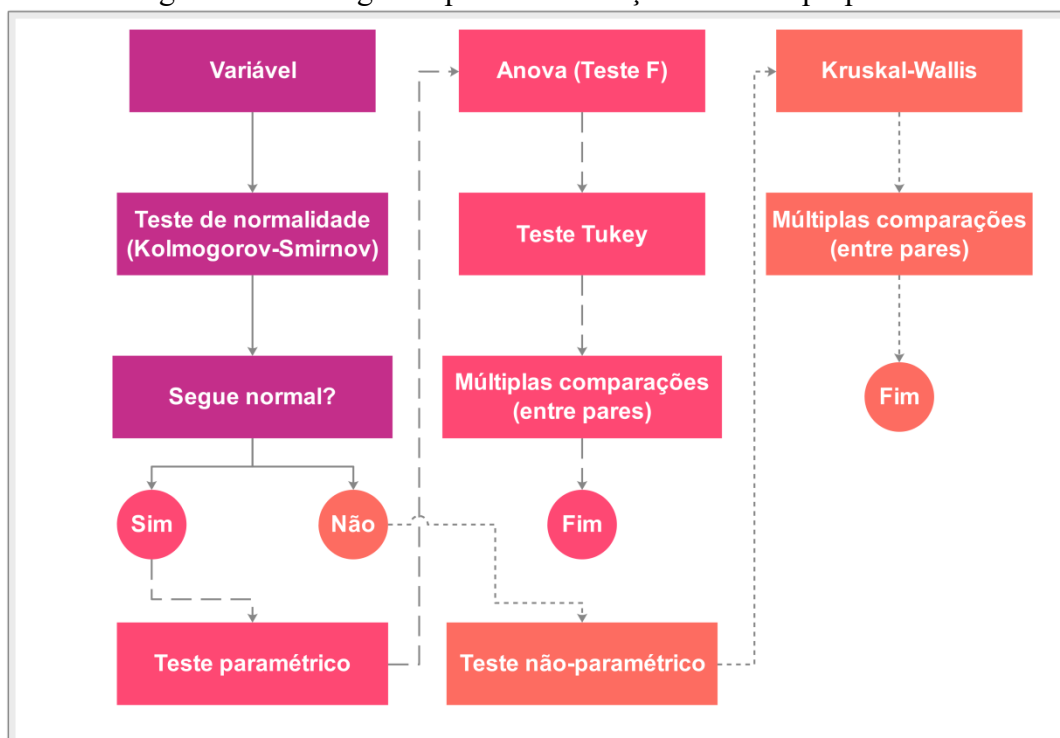
A fase de caracterização dos padrões de viagens dos usuários consistirá na validação das hipóteses propostas. Essa fase será fundamentada em duas etapas: testes estatísticos e conclusões sobre o fenômeno. Na etapa de testes estatísticos, as hipóteses serão investigadas através da realização dos testes estatísticos. E na etapa de conclusões

sobre o fenômeno, serão elaboradas conclusões baseando-se nos resultados da etapa anterior que investigará as hipóteses propostas anteriormente.

Os testes estatísticos adotados serão o ANOVA, Kolmogorov-Smirnov e o Teste Tukey para efetuar a análise sobre o comportamento de viagens dos indivíduos entre os dias da semana. O nível de significância adotado para realização dos testes será de 95% e a variável analisada será o número diário de viagens dos usuários por dia da semana.

Inicialmente, será feita uma análise descritiva da variável para cada dia da semana, através de gráficos *box plots*. Em seguida, antes de realizar o teste F ANOVA, será verificada a normalidade da variável número diário de viagens por usuário, através do teste de normalidade Kolmogorov-Smirnov. Se o teste Kolmogorov-Smirnov rejeitar a hipótese de normalidade, será aplicado o teste de Kruskal-Wallis. Esse teste não paramétrico é análogo ao ANOVA e não leva em consideração a normalidade. Além disso, diferente do teste ANOVA, a hipótese nula afirma que as amostras são de populações idênticas. Essa explicação pode ser mais bem compreendida na Figura 17, que ilustra a etapa de testes estatísticos dessa fase do estudo.

Figura 17 - Fluxograma para identificação do teste apropriado



Fonte: Elaborado pelo autor.

O objetivo do ANOVA nessa etapa é testar se há alguma diferença significativa entre as médias de viagens de dois ou mais grupos, de modo que a hipótese nula é representada por  $H_0: \mu_{seg} = \mu_{ter} = \mu_{quar} = \mu_{quin} = \mu_{sex} = \mu_{sáb} = \mu_{dom}$ , validando ou rejeitando a primeira hipótese do estudo, a de que há diferença no comportamento de viagens dos indivíduos conforme os dias da semana.

No entanto, se apenas uma dessas médias diferir das demais, isso implicará na rejeição da hipótese nula e aceitação da hipótese alternativa. Além disso, o teste ANOVA não especifica qual das médias que diferiu das demais, então, em seguida será realizado o teste Tukey (caso haja rejeição da hipótese nula).

Nesse teste, as diferenças entre grupos são chamadas de diferenças em pares. O método Tukey testará a maior diferença entre os pares, utilizando o parâmetro estatístico “ $q$ ” para determinar se há diferença significativa entre os pares.

Esse parâmetro é obtido, subtraindo-se a menor da maior média e dividindo esse produto pelo erro padrão geral da média do grupo (BENJAMINI e HOCHBERG, 1995). O valor “ $q$ ” é então confrontado em uma tabela de valores “ $q$ ” para averiguar se o valor “ $q$ ” de um par específico excede o valor “ $q$ ” crítico necessário para determinar a significância estatística. Caso o valor “ $q$ ” atinja ou exceda o valor crítico, a diferença desse par é estatisticamente significativa. Porém, devido à facilidade de interpretação, será analisado o “valor- $p$ ”. Se o “valor- $p$ ” do par em análise for maior do que o “ $\alpha$ ” adotado, no caso 0,05, considera-se que não há diferença significativa entre o par.

O teste Tukey também identificará quais grupos são diferentes dos demais, sendo útil assim para investigar as demais hipóteses do estudo. O teste ANOVA, por sua vez, é incapaz de fornecer essa informação, por isso, foi utilizado apenas para investigar a primeira hipótese. Esses testes foram realizados por meio do software de programação computacional R.

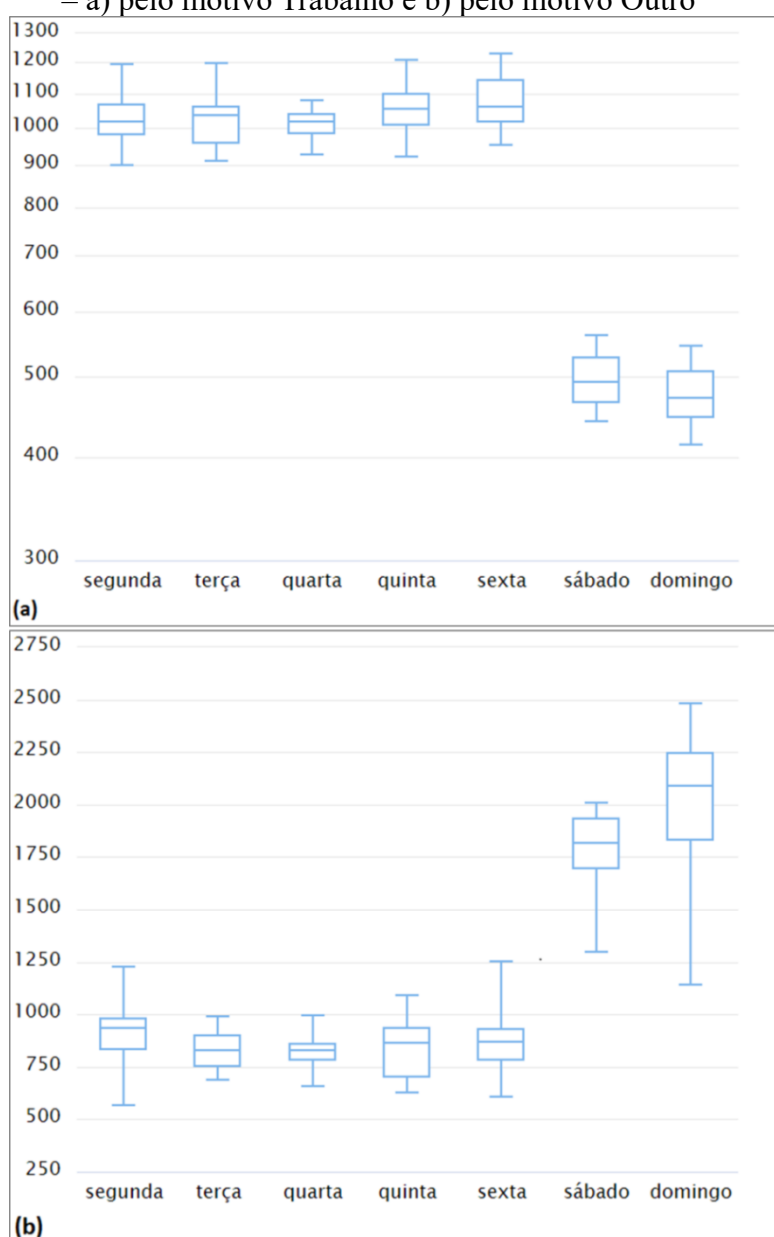
### **3.4 Resultados e discussões**

Nesta seção são apresentados os resultados acerca da análise da variabilidade de viagens dos usuários do sistema de transportes de Fortaleza através de dados de redes sociais. O resultado está organizado em diferentes níveis de detalhes, analisando o comportamento entre os diferentes dias da semana ao longo dos meses considerados no estudo.

### 3.4.1 Caracterização dos padrões de viagens em Fortaleza

Antes de proceder aos testes estatísticos, executou-se uma análise de estatística descritiva da variável número de viagens por motivo, através dos gráficos *box plot* da variável em análise presentes na Figura 18. Na Figura 18a, encontram-se os gráficos referentes ao número de viagens motivo trabalho por usuário para cada dia da semana, e na Figura 18b, número de viagens pelo motivo Outro.

Figura 18 - *Box plots* do número de viagens por usuário por dia da semana em Fortaleza – a) pelo motivo Trabalho e b) pelo motivo Outro



Fonte: Elaborado pelo autor.

Ao visualizar os gráficos da Figura 18, observa-se que tanto a Figura 18a como a Figura 18b apresentam padrões diferentes entre os dias da semana, como esperado, corroborando a hipótese de que há diferença no comportamento de viagens dos usuários conforme o dia da semana em Fortaleza. Essa diferença é muito maior comparando os dias úteis da semana com os finais de semana do que comparando os dias úteis da semana entre si e isso ocorre tanto para viagens por motivo Trabalho, quanto para viagens pelo motivo Outro.

Quanto à hipótese de que o comportamento de viagens dos indivíduos na sexta-feira é parecido ao dos finais de semana, não é possível afirmar tal afirmação, uma vez que na sexta-feira o comportamento dos usuários se assemelha bastante com o dos demais dias da semana sendo que na sexta-feira os usuários realizam um número ligeiramente maior de viagens do que nos demais dias da semana. O mesmo pode ser observado em viagens pelo motivo Outro. No entanto, o resultado referente a esse dia, não deve ser comparado com o resultado dos finais de semana, que são muito diferentes quando comparados com qualquer dia útil.

De posse disso, pode-se afirmar que há diferença no comportamento de viagens dos indivíduos entre os dias da semana, sendo essa diferença mais significativa comparando-se os dias úteis da semana com os finais de semana, não havendo tal diferença entre os dias úteis. Percebe-se, através dos resultados apresentados nos gráficos, que há evidências empíricas para validar a hipótese de que o comportamento de viagens varia conforme os dias da semana, mas não é possível validar as demais hipóteses, principalmente aquela que afirma que há diferença entre a sexta-feira e os demais dias úteis. Portanto, se faz necessário realizar testes paramétricos para validar as demais hipóteses.

Quanto à variabilidade dos motivos Trabalho e Outro, observa-se através da análise dos primeiros e terceiros quartis dos gráficos de *box plot*, que para viagens motivo Trabalho há um grau de variação maior do que para viagens pelo motivo Outro, contrariando a hipótese inicialmente proposta, qual seja, de que a variabilidade comportamental dos usuários em viagens pelo motivo Outro era maior do que em viagens por motivo Trabalho. Além disso, a variabilidade das viagens pelo motivo Outro é notadamente maior nos finais de semana do que na semana, corroborando com a hipótese proposta no começo do estudo. Se comparada com a variabilidade de viagens por motivo Trabalho, nota-se, no entanto, que a variabilidade de viagens pelo motivo Outro é menor em dias úteis e no final de semana é tão grande quanto a de viagens pelo motivo Trabalho.

Isso só ocorre durante os dias da semana, nos finais de semana, a variabilidade de ambos os tipos de viagem são equiparáveis. Isso possivelmente se deve à variabilidade nas preferências dos indivíduos, uma vez que nos fins de semana, há pessoas que preferem ficar em casa ao invés de sair para realizar outras atividades. Além disso, comparando os dias da semana com os finais de semana, observa-se através da posição da mediana que nos finais de semana se realizam um número muito menor de viagens por motivo Trabalho, que nos dias úteis da semana e se realizam significativamente mais viagens pelo motivo Outro. Essa observação fenomenológica não só corrobora a hipótese estabelecida no começo do estudo, como atesta a confiabilidade da base de dados obtida, uma vez que isso é algo recorrente, pois aos finais de semana, a maioria das pessoas está mais livre para realizar outras atividades que não seja trabalhar e, supostamente, tendem a realizar mais atividades do tipo Outro.

### **3.4.2 Testes estatísticos**

Concluída a análise preliminar, seguiu-se para os testes estatísticos de fato para poder validar as hipóteses propostas anteriormente. O primeiro teste realizado foi o teste de normalidade Kolmogorov-Smirnov, da variável número de viagens por dia de cada usuário. Tal teste mostrou que a variável segue normal para todos os conjuntos de dias da semana e de meses. Por isso, foi possível utilizar o teste estatístico de análise de variância ANOVA.

Para o teste ANOVA utilizou-se dados de 31 semanas (217 dias). Os dados referentes a cada dia da semana foram agrupados de maneira que cada grupo compreendesse dados de um dia da semana diferente, totalizando sete grupos, sendo cada grupo referente a um dia da semana específico. Esses sete grupos de semanas foram divididos em dois grupos classificados de acordo com o motivo de viagens (Trabalho e Outro), e após isso, foi aplicado o teste ANOVA para cada motivo de viagem distinto.

Os resultados ensejam indícios suficientes para a rejeição da hipótese nula, ou seja, a média de no mínimo um grupo de dias da semana é diferente dos demais. As Tabelas 2 e 3 exibem os parâmetros utilizados na análise, como o F e o F crítico, além do valor-p. Os parâmetros como o valor F e o F crítico foram empregados como critérios para a rejeição da hipótese nula. Como o F calculado é maior do que o F crítico, em ambos os casos (motivos) há indícios suficientes para a rejeição da hipótese nula. Além dos



parâmetros F, analisou-se o valor-p, dessa forma, o valor-p foi de  $\alpha < 0.05$ , assim sendo, há indícios suficientes de que existe pelo menos uma média com diferença significativa das demais com nível  $\alpha$  de significância adotado.

Tabela 2 – Anova (pelo motivo Trabalho)

	<b>Sum Sq</b>	<b>Df</b>	<b>Valor-F</b>	<b>Pr (&gt;F)</b>
Dia	13688804	6	746.074	1.28E-138
Resíduos	642172.4	210	NA	NA

Fonte: Elaborado pelo autor.

Tabela 3 – Anova (pelo motivo Outro)

	<b>Sum Sq</b>	<b>Df</b>	<b>Valor-F</b>	<b>Pr(&gt;F)</b>
Dia	50704287	6	285.325	4.95E-98
Resíduos	6219750	210	NA	NA

Fonte: Elaborado pelo autor.

Em seguida, realizou-se o teste Tukey (HSD – Diferença Significativa de Tukey) para determinar se houve diferença expressiva entre todos os pares possíveis de dias da semana. Esse teste para ambos os motivos, indica diferença significativa entre alguns dias, com  $P \text{ adj} < 0.05$ . A Tabela 4 apresenta os pares entre os dias para o motivo Trabalho, e a Tabela 5 para o motivo Outro. A partir das Figuras 19 e 20 foi possível visualizar graficamente a diferença observada.

Tabela 4 - Teste Tukey (pelo motivo Trabalho)

Dias (pares)	diff	lwr	upr	P adj
quarta-domingo	538.7742	496.9574	580.591	0
quinta-domingo	577.3226	535.5058	619.1394	0
sábado-domingo	18.25806	-23.5587	60.07486	0.85142
segunda-domingo	549.7742	507.9574	591.591	0
sexta-domingo	601.6129	559.7961	643.4297	0
terça-domingo	549.129	507.3122	590.9458	0
quinta-quarta	38.54839	-3.26841	80.36518	0.092552
sábado-quarta	-520.516	-562.333	-478.699	0
segunda-quarta	11	-30.8168	52.81679	0.986373
sexta-quarta	62.83871	21.02192	104.6555	0.000251
terça-quarta	10.35484	-31.462	52.17163	0.99009
sábado-quinta	-559.065	-600.881	-517.248	0
segunda-quinta	-27.5484	-69.3652	14.26841	0.442513
sexta-quinta	24.29032	-17.5265	66.10712	0.597164
terça-quinta	-28.1935	-70.0103	13.62324	0.413202
segunda-sábado	531.5161	489.6993	573.3329	0
sexta-sábado	583.3548	541.538	625.1716	0
terça-sábado	530.871	489.0542	572.6878	0
sexta-segunda	51.83871	10.02192	93.6555	0.005226
terça-segunda	-0.64516	-42.462	41.17163	1
terça-sexta	-52.4839	-94.3007	-10.6671	0.00444

Fonte: Elaborado pelo autor.

Figura 19 - Representação gráfica do teste Tukey (pelo motivo Trabalho)

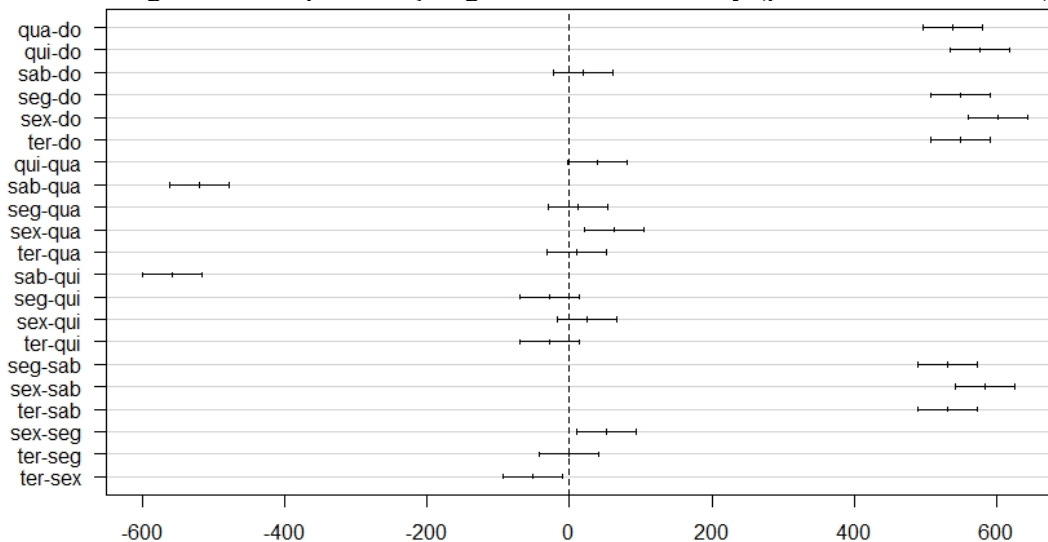
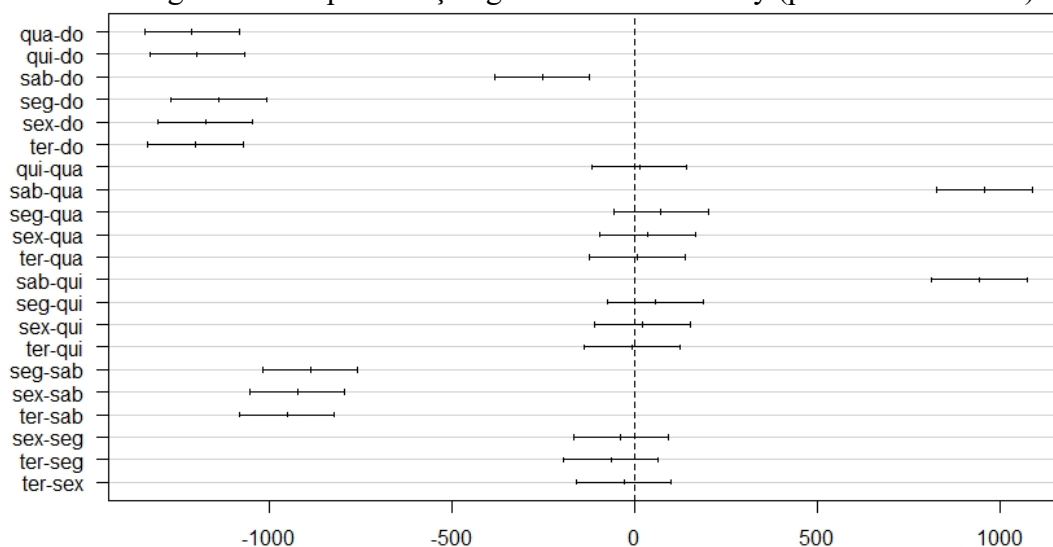


Tabela 5 - Teste Tukey (pelo motivo Outro)

Dias (pares)	diff	lwr	upr	p adj
<b>quarta-domingo</b>	-1210.9	-1341.04	-1080.76	0
quinta-domingo	-1196.35	-1326.5	-1066.21	0
<b>sábado-domingo</b>	-251.419	-381.56	-121.279	0.0000006
segunda-domingo	-1137.71	-1267.85	-1007.57	0
<b>sexta-domingo</b>	-1174.03	-1304.17	-1043.89	0
terça-domingo	-1202	-1332.14	-1071.86	0
<b>quinta-quarta</b>	14.54839	-115.592	144.6886	0.9998891
<b>sábado-quarta</b>	959.4839	829.3437	1089.624	0
<b>segunda-quarta</b>	73.19355	-56.9466	203.3337	0.6338644
<b>sexta-quarta</b>	36.87097	-93.2692	167.0111	0.9800012
<b>terça-quarta</b>	8.903226	-121.237	139.0434	0.9999939
<b>sábado-quinta</b>	944.9355	814.7953	1075.076	0
<b>segunda-quinta</b>	58.64516	-71.495	188.7853	0.8314791
<b>sexta-quinta</b>	22.32258	-107.818	152.4628	0.9986876
<b>terça-quinta</b>	-5.64516	-135.785	124.495	0.9999996
<b>segunda-sábado</b>	-886.29	-1016.43	-756.15	0
<b>sexta-sábado</b>	-922.613	-1052.75	-792.473	0
<b>terça-sábado</b>	-950.581	-1080.72	-820.44	0
<b>sexta-segunda</b>	-36.3226	-166.463	93.81759	0.9814786
<b>terça-segunda</b>	-64.2903	-194.431	65.84985	0.7618522
<b>terça-sexta</b>	-27.9677	-158.108	102.1724	0.9953873

Fonte: Elaborado pelo autor.

Figura 20 - Representação gráfica do teste Tukey (pelo motivo Outro)



Fonte: Elaborado pelo autor.

Analisando-se os resultados a partir do teste Tukey, como esperado, a quantidade de viagens para o domingo difere dos dias da semana para ambos os motivos.

A partir da análise, pode-se perceber que em relação ao motivo Outro a quantidade de viagens do domingo difere do sábado, talvez haja indícios de que as atividades realizadas por motivo lazer não sejam as mesmas realizadas aos domingos. Segundo Ahas *et al.* (2010) e Allstrom *et al.* (2017) o sábado é influenciado pelos impactos dos dias úteis ou pela necessidade de realizar compras. O sábado à tarde é um horário típico para realizar atividades como compras e lazer.

Por outro lado, ao se comparar a média de viagens por motivo Trabalho entre sábado e domingo, não foram observadas variações significativas. Isso pode ser explicado pela baixa quantidade de viagens por esse motivo no final de semana, sendo mais comum esse tipo de motivo durante a semana. Segundo Raux *et al.* (2016) aos domingos há uma queda no número de viagens, fazendo, definitivamente, com que esse dia pareça como um dia diferente dos outros dias da semana. Isso indica uma especificidade do sábado e domingo em sequência de atividades quando comparados com o restante dos dias úteis.

Em contrapartida, ao se observar a variabilidade de viagens por motivo Outro durante os dias úteis da semana, pode-se verificar que não houve variação no número de viagens entre os dias, se assemelhando ao comportamento observado nos gráficos *box plot* que indicavam que não havia grande variação no comportamento dos indivíduos em viagens motivo Outro entre os dias úteis da semana. Em atividades de lazer, como as que ocorrem em lotes considerados como Outro nesse estudo: esportes, religião e estudo, as magnitudes da variabilidade no número de viagens tendem a não variar significativamente.

Por outro lado, as viagens por motivo Trabalho apresentaram baixa variabilidade, resultado esse, semelhante ao discutido em (TARIGAN e KITAMURA, 2009). O teste tukey mostrou que apenas a sexta-feira difere dos demais dias úteis da semana, validando a hipótese proposta anteriormente de que na sexta-feira o comportamento dos usuários é atípico se comparado ao dos outros dias úteis da semana. De fato, a sexta-feira apresentou uma quantidade maior de viagens por motivo Trabalho, Figura 18a. Esse resultado é semelhante ao apresentado em Dharmowijoyo *et al.* (2015), em que foi observado que a sexta-feira possui um alto número de viagens em comparação com os outros dias úteis.

De modo geral, os resultados revelaram uma variabilidade de viagens significativa, com sequências de viagens semelhantes, apenas diferindo alguns dias, confirmando alguns resultados obtidos em trabalhos anteriores como os de (DHARMOWIJOYO *et al.*, 2015; AHAS *et al.*, 2010; ALLSTROM *et al.*, 2017; RAUX

*et al.*, 2016). Para padrões nos dias úteis da semana, a variabilidade da sexta-feira é bem maior que as dos demais dias úteis da semana para ambos os motivos, validando a hipótese proposta. Uma revelação inesperada foi que o mesmo ocorre na quinta-feira. Esses resultados revelam que existe uma espécie de “transição” nos padrões de variabilidade de viagens entre o começo da semana e os finais de semana.

### **3.5 Considerações finais**

As análises dos dados obtidos através do *Instagram* e do *Twitter* forneceram um estudo acerca da variabilidade de viagens. Essencialmente, foi realizada uma divisão dos dados em dois grupos de viagens por motivo, Trabalho e Outro. Em seguida, foram aplicados testes estatísticos com a finalidade de compreender que tipo de distribuição a variável viagem seguia e qual o método que melhor se adaptaria a análise. Neste capítulo, foi visto que a variabilidade pode ser medida levando em consideração os diferentes dias da semana, mais especificamente, a variabilidade foi medida como um indicador de viagens do tipo interpessoal.

Os resultados deste estudo forneceram evidências fenomenológicas que conduziram ao entendimento da variabilidade no comportamento diário de viagens para os diferentes motivos considerados. Com isso, se pode obter não apenas conclusões importantes sobre o fenômeno, como constatar o fato de se haver descoberto um período de transição entre os dias da semana e o final de semana na quinta e na sexta. Além disso, teceram-se considerações que atestam a confiabilidade e a representatividade dos dados e do método proposto para o estudo, como o fato de a frequência de viagens por motivo Trabalho ter sido maior do que pelo motivo Outro durante a semana, e o contrário, nos finais de semana.

Adicionalmente, pode-se observar a influência dos dias da semana no comportamento dos indivíduos quando se verificou que a sexta-feira tende a ter uma média maior de viagens e que isso se deve provavelmente a proximidade do final de semana e a sua influência no comportamento dos indivíduos nesse dia. Finalmente, acredita-se que esse estudo forneceu informações satisfatórias sobre o comportamento da variabilidade de viagens a partir de dados de redes sociais, e como essas viagens podem ser empregadas para fornecer padrões a respeito da variabilidade. Além disso, com base na compreensão do fenômeno, vale salientar que as análises acerca da variabilidade são

provenientes de viagens realizadas em sua grande maioria por indivíduos de média e alta renda, o que sob esse aspecto, torna a amostra enviesada.

## 4 ANÁLISE COMPARATIVA ENTRE UMA MATRIZ OD OBTIDA ATRAVÉS DE USUÁRIOS DE REDES SOCIAIS E UMA MATRIZ OD EXTERNA E POSSÍVEIS APLICAÇÕES

### 4.1 Introdução

A Matriz OD é um dos principais *inputs* para a grande maioria dos projetos e modelos de análise de demanda por transportes. Ela descreve os padrões da mobilidade urbano e é considerada o primeiro passo a ser realizado antes do emprego do modelo sequencial (MOHAMED *et al.*, 2016). Uma das formas mais comuns para sua obtenção é através da pesquisa OD domiciliar (EGU e BONNEL, 2020). Esse tipo de pesquisa é cuidadosamente projetado com precisão, e apoiada em procedimentos de amostragem cuidadosos, cujo os resultados são combinados aos aspectos socioeconômicos. Apesar das vantagens, ainda é considerado um método caro, com quedas nas taxas de respostas (Atrostic *et al.*, 1999; Groves, 2006) e com baixa frequência de repetição, sendo esse por último a sua maior desvantagem.

Com o surgimento das tecnologias de informação, houve uma mudança na forma da obtenção da matriz OD. Em particular, a disponibilidade dos *smartphones* e a disseminação das redes sociais geraram determinados conjuntos de dados que quando estruturados e analisados, fornecem possibilidades sem precedentes acerca da análise dos padrões de mobilidade (MIRZAEI e WANG, 2020). Consequentemente, a utilização dos dados de redes sociais para a compreensão da mobilidade possui algumas fragilidades em comparação aos métodos tradicionais de coleta, mais precisamente a pesquisa OD domiciliar. Um dos condicionantes é a importância dos processos de filtragem e pré-processamento dos dados, para que desse modo, obtenha-se um conjunto de dados com informações confiáveis acerca da mobilidade urbana. Outro fator sobre a limitação desse tipo de dado é a baixa resolução temporal.

Em se tratando da mobilidade urbana, a possibilidade de gerar matrizes OD é fundamental para a previsão da demanda por viagens, modelagem por transportes e otimização do uso da rede (GAO *et al.*, 2014). Entretanto, as matrizes OD advindas das redes sociais necessitam passar por um processo de comparação com estudos externos (matrizes OD extraídas de pesquisas domiciliares) antes de serem empregadas para

análise de transportes, especialmente considerando que elas servirão de entrada para modelos de transporte.

Alguns trabalhos desenvolveram diversas técnicas para converter esses dados em viagens, porém poucos tentaram realizar sua “validação”, ou seja, a comparação entre as viagens advindas das redes sociais com dados de uma fonte confiável, a fim de obter uma opinião mais clara acerca do seu potencial. Uma matriz OD das redes sociais confiável, essa serviria de base complementar a pesquisa OD domiciliar, além de ser do interesse de órgãos governamentais e empresas de consultoria.

Desse modo, este capítulo tem como objetivo principal realizar uma análise comparativa e sistemática a partir de duas matrizes OD, a saber: uma matriz obtida a partir das redes sociais (Capítulo 2) e uma matriz OD estimada por meio do modelo integrado de uso do solo e transportes TRANUS. Para tal finalidade, apresentam-se como objetivos específicos: i) comparar a matriz obtida das redes sociais com uma matriz OD externa e ii) Identificar a utilidade desse tipo de matriz OD para o planejamento urbano de transportes (análise de aplicabilidade). Este capítulo está subdividido em cinco seções, a primeira delas é introdução, na qual o trabalho é apresentado. Na seção 4.2 apresenta-se uma revisão da literatura com os principais aspectos sobre a matriz OD das redes sociais, sendo eles: matriz OD coletada a partir de dados das redes sociais, potenciais e limitações e aplicabilidade de dados de redes sociais no planejamento de transporte. Na seção 4.3 encontram-se os procedimentos metodológicos necessários para a comparação entre as matrizes OD. Em seguida, na seção 4.4, são apresentados os resultados da comparação e possíveis aplicabilidades da matriz OD das redes sociais. Por fim, na seção 4.5, encontram-se as conclusões, onde são apresentadas as principais contribuições do trabalho, bem como suas limitações.

## **4.2 Revisão Bibliográfica**

### ***4.2.1 Matriz OD coletada a partir de dados das redes sociais***

O tráfego urbano está se tornando cada vez mais complexo, caracterizado pelo aumento do número de viagens e uso mais intenso de veículos motorizados. Diante dessa



complexidade, as fontes tradicionais de dados precisam ser complementadas com novos dados que possuam uma maior resolução espacial e temporal. Nesse caso, as diferentes possibilidades de dados podem servir de base para descrever padrões e características de diferentes indivíduos no que se refere a mobilidade urbana.

Algumas investigações utilizaram dados de redes sociais para a construção de matriz OD. Focando no *Twitter* como ferramenta para a construção de matriz OD, os autores Jin *et al.* (2014) estimaram uma matriz OD em Chicago baseada no modelo gravitacional duplamente restrito, com uso dos dados advindos do *Foursquare*. A metodologia proposta consistiu na calibração e avaliação da matriz OD através de uma comparação com a matriz OD gerada pelo método gravitacional de restrição única, como também através de uma matriz fornecida pelo planejamento metropolitano local. Os resultados mostraram uma melhoria significativa na redução de erros de estimativa causado pelo viés de amostragem baseado no modelo gravitacional.

Yang *et al.* (2015) estimaram uma matriz OD a partir de *check-in* de usuários do *Foursquare* na área urbana de Chicago. A matriz OD foi validada por meio de uma comparação entre os diferentes comprimentos de viagens, mapas de calor da produção e dos destinos de viagens confrontados com dados reais. Os resultados mostraram-se semelhantes quando comparados com dados reais, desse modo, foram considerados promissores quando se refere a compreensão dos diferentes padrões de mobilidade.

Kheiri *et al.* (2015) examinou a eficiência dos *check-ins* advindos do *Foursquare*, aplicados a obtenção de uma matriz OD do distrito de Manhattan, sendo esse um dos mais populosos que compõem a cidade de New York. Os dados utilizados nessa pesquisa foram provenientes de 1083 usuários, coletados em um intervalo de dez meses. A metodologia consistiu na aplicação de três métodos de estimação de viagens, sendo eles: *Radiation model* (fluxo de viagens de uma zona  $i$  para uma zona  $j$ ), *Rank-based model* (a distância sendo um fator chave para que haja o deslocamento) e *Poupolation-weighted opportunities model* (modelo de oportunidades ponderado pela população, onde a probabilidade de viajar de  $i$  para  $j$  é proporcional à destinos de  $j$ ). Após a obtenção do fator de contribuição de cada modelo, eles foram comparados com a pesquisa MTA (*Metropolitan Transportation Authority*). O modelo que melhor se aproximou da realidade foi o *Rank-based model*, esse possuiu um excelente desempenho ao utilizar dados LBSN do que os outros modelos.

Hu e Jin (2018) estimaram uma matriz OD na ilha de Manhattan em Nova York. Os dados empregados foram coletados a partir do *Foursquare*. Eles utilizaram coeficientes de correlação para comparar os padrões de *check-ins* entre origens e destino. Os diferentes pares OD foram integrados à função de atrito de modelo gravitacional. Os resultados mostraram que os padrões de fluxos OD por hora alcançaram um erro absoluto mínimo e o  $R^2$  de regressão foi de aproximadamente 0,78. Os resultados apresentaram o potencial acerca desse tipo de dado para a estimação da matriz OD e análise da demanda de viagens urbanas.

Os autores Arjona e Palomares (2019) empregaram *tweets* referentes a área metropolitana de Madrid ao longo de dois anos. Eles realizaram a coleta, limpeza e estruturação dos dados, com a finalidade de obtenção da matriz OD. Após a coleta, essa matriz OD foi comparada com os dados oficiais referentes a uma matriz sintética referente ao ano de 2014. Os resultados desta comparação apresentaram que o nível de precisão foi adequado e eficiente.

Os resultados alcançados nesses estudos mostraram o potencial das redes sociais na estimação de matrizes OD. No entanto, são necessários análises mais detalhadas a respeito dessas matrizes para a sua real aplicabilidade em planejamento de transportes urbanos.

#### **4.2.2 Potenciais e limitações dos dados advindos das redes sociais**

Convencionalmente, as matrizes OD são derivadas por um processo complexo, que envolve diferentes métodos de coleta. Isso inclui alguns métodos, como a pesquisa OD domiciliar (Pires *et al.*, 1997) e a pesquisa a partir de contagens volumétricas (ROBILLARD, 1975; CASCETTA, 1984; HAZELTON, 2008). Essas metodologias são consideradas confiáveis e amplamente difundidas. Recentemente, as viagens obtidas através das redes sociais vêm sendo aplicadas em diferentes áreas do planejamento de transportes, no que se refere ao desenvolvimento de matrizes OD. Essas matrizes têm o potencial de servirem de base complementar aos diversos tipos de coleta convencionais. Além disso, elas possuem uma cobertura geográfica aceitável e um baixo custo de obtenção. A obtenção e aplicação desses dados em planejamento de transportes é relativamente nova, possuindo vantagens e desvantagens em comparação com as

abordagens tradicionais. Quando empregados para obtenção de uma matriz OD há um risco de vieses que precisam ser avaliados. Outro problema é a dificuldade da identificação de viagens curtas, assim como o reconhecimento do modo de transporte utilizado no deslocamento (RASHIDI *et al.*, 2017). Tais problemas, contrastam com as abordagens tradicionais, que permitem questionar diretamente ao entrevistado qual o modo de transporte empregado para se deslocar.

Esses desafios deixam evidente que, embora haja benefícios do uso desses dados para a obtenção da matriz OD, esses não são equiparáveis às abordagens tradicionais, ainda que a matriz OD das redes sociais forneça informações a respeito da mobilidade urbana, essa possui suas limitações. Uma outra desvantagem associada ao uso das redes sociais, refere-se à baixa consistência (SUN e LI, 2015). Um outro viés refere-se ao fato de que os dados de redes sociais refletem os deslocamentos de indivíduos pertencentes a uma renda mais elevada.

Os autores Pourebrahim *et al.* (2019) afirmaram que os dados do *Twitter* representam apenas os fluxos de viagens de pessoas de renda mais elevada. Daniele *et al.* (2015) afirmam que dados de redes sociais não representam todas as classes socioeconômicas, tendo como base a suposição de que apenas uma parte da população possui acesso ao smartphone e à internet. Li *et al.* (2013) empregaram dados georreferenciados do *Twitter* e do *Flickr* com a finalidade de compreender padrões espaciais e temporais dessas fontes de dados a nível nacional nos Estados Unidos. Variáveis socioeconômicas foram aplicadas ao modelo de regressão e os resultados indicam que pessoas com um perfil socioeconômico mais elevado possuem uma maior atividade nas redes sociais. Os autores Robertson e Feick (2016) observaram que esses dados não possuem o potencial de descrever áreas de extrema pobreza.

Ademais, a quantidade dos dados a partir de redes sociais está em grande parte associada à propriedade de um smartphone e a conexão de uma internet móvel. Sloan e Quan-haase (2017) afirmaram que o *Twitter* é utilizado apenas por uma parcela da população e que parte dos seus usuários não estão com o receptor GPS ativo. Segundo Jiang *et al.* (2019), nem todos os usuários do *Twitter* compartilham a sua localização, resultando assim em um viés populacional que pode acarretar em uma superestimação ou subestimação do modelo. Segundo Jiang e Thill (2015), esses vieses precisam ser considerados ao empregar informações geográficas voluntárias, ou seja, dados passivos. Por outro lado, mesmo que esses dados sejam aceitáveis em termos de quantidade, a

representatividade a respeito da informação fornecida foi reconhecida como um problema. Segundo Kung *et al.* (2014) nem todas as pessoas usam redes sociais. Omri (2016) realizou uma pesquisa em nível nacional a respeito da representatividade dos dados advindos do *Twitter* nos Estados Unidos. Ele concluiu que esses dados podem ser representativos para um estado, mas não necessariamente para outro. Isso sugere que, independentemente de quão representativos sejam os dados do *Twitter* em nível nacional, a representatividade em nível estadual pode ser completamente diferente. Isso significa que, para alguns estados, os *tweets* podem ser representativos da população, enquanto, para outros, os *tweets* podem ser completamente insignificantes. Quercia *et al.* (2015) afirmam que os usuários das redes sociais não são necessariamente uma amostra representativa e sim que apenas constituem uma parte da população cujo uso tende a estar alinhado a um interesse específico. Conforme evidenciado pelos trabalhos citados, apesar desses dados possuírem as suas limitações, são considerados um subproduto valioso para o estudo dos padrões de deslocamentos.

Embora os estudos mencionados anteriormente apresentem o potencial dos dados das redes sociais, deve-se mencionar as suas vantagens e desvantagens. É necessário que haja mais pesquisas para a compreensão do potencial desse tipo de dado. Apesar desses estudos terem sido realizados na compreensão dos padrões de mobilidade, é essencial mencionar o erro introduzido nos padrões devido à baixa representatividade e vieses presentes nos dados.

#### **4.2.3 Aplicabilidade de dados de redes sociais no planejamento de transporte**

Nos últimos anos, diversos trabalhos relacionados aos dados de redes sociais foram publicados. A maioria dessas pesquisas apresentaram o potencial destes dados na área do planejamento de transportes. Segundo Nikolaidou e Papaioannou (2018) existem três características importantes das redes sociais que interessam aos analistas de transportes. A primeira se refere ao conteúdo postado, geralmente a eventos ou ações vivenciadas por indivíduos. A segunda aborda que cada evento está relacionado a um determinado intervalo de tempo, ou seja, os usuários tendem a postar informações sobre o que aconteceu ou vai acontecer em um dado local. Por fim, cada evento é tido como algo importante para esses usuários. Segundo Gal-Tzur *et al.* (2014), as informações

publicadas nas redes sociais a respeito dos diversos acontecimentos na área de transporte são divididas em três tópicos:

- Informações sobre a necessidade de se deslocar;
- Obtenção de um evento que pode causar impacto na mobilidade, como congestionamento, acidente, etc;
- Opiniões dos usuários acerca da qualidade de um dado serviço de transporte.

A coleta dos dados das redes sociais como o *Twitter*, pode ser realizada em tempo real. Esses dados quando tratados, podem ser empregados para a detecção de acidentes, informações sobre tráfego, dentre outras (CHANIoTAKIS e ANTONIOU, 2015). De modo geral, os dados das redes sociais vêm sendo aplicados a área de transportes, principalmente quando se refere a compreensão da mobilidade urbana, eventos no trânsito e obtenção de matrizes OD.

#### 4.2.3.1 Detecção de Eventos

Alguns trabalhos originados a partir das redes sociais foram empregados para a detecção de eventos no trânsito. Wanichayapong *et al.* (2011) desenvolveram um algoritmo capaz de obter informações a respeito do tráfego em tempo real. As informações foram extraídas através de uma análise textual dos *tweets* e classificadas em duas categorias: ponto e link. Primeiramente, as informações do ponto foram associadas a um local (por exemplo, um acidente de carro em uma interseção), enquanto as informações do link foram associadas aos pontos de partida e chegada em uma rodovia (por exemplo, o congestionamento entre duas interseções).

Gu *et al.* (2016) aplicaram uma metodologia de análise de texto, capaz de minerar *tweets* e obter informações relacionadas a incidentes que ocorrem no trânsito de forma barata. A metodologia foi empregada em duas regiões metropolitanas de Pittsburgh e Filadélfia. Os resultados forneceram informações razoáveis sobre determinadas vias arteriais. Eles descobriram que há mais informações sobre acidentes de trânsito nos finais de semana do que durante a semana. Além disso, observou-se que há mais ocorrências relatadas pelos usuários durante o dia.

Zhang *et al.* (2016) verificaram a potencialidade dos *tweets* em representar determinados congestionamentos na rede viária dentro em um dado espaço de tempo. Eles combinaram um conjunto de *tweets* e dados a respeito do tráfego na Virgínia do Norte,

assim, foi obtida a correlação entre dois bancos de dados. Os resultados comprovaram o potencial dos *tweets* na detecção dos diversos picos ao longo do dia. Eles verificaram ainda a correlação entre a quantidade de *tweets* e o aumento do tráfego em algumas vias, e concluíram que as principais atividades do *Twitter* se relacionavam ao tráfego intenso.

Ni *et al.* (2016) desenvolveram uma abordagem sistemática com a finalidade de prever o volume de passageiros em uma estação de metrô, através da utilização dos dados advindos do *Twitter*. O estudo foi realizado em New York e puderam observar que existe uma correlação entre o fluxo de passageiros e os *check-ins*. A metodologia consistiu em um algoritmo capaz de detectar determinados eventos no trânsito a partir da leitura de hashtag. Além disso, eles propuseram uma abordagem baseada em otimização paramétrica, conhecida como função de perda híbrida. Os resultados mostraram que houve um melhor desempenho do modelo ao empregar os dados.

Paule *et al.* (2019) propuseram um método de inferência que utiliza uma abordagem de classificação a partir de *tweets* geolocalizados, com a finalidade de obter incidentes no trânsito. O estudo foi realizado em Chicago, New York. Os resultados apresentaram que o método de geolocalização dos *check-ins* pode ser útil para a compreensão e mapeamento de incidentes no trânsito.

Xu *et al.* (2019) desenvolveram um método capaz de minerar e associar eventos ocorridos no trânsito. O estudo empregou dados do *Twitter* e foi realizado em Toronto, Canadá. Os *tweets* empregados foram treinados a partir de regras de associação. Diversas regras foram criadas com a utilização de aprendizagem de máquina. Essas regras se dividiram em: trânsito, acidentes, obras na rodovia e condições climáticas. Após a definição das regras, eles compararam os eventos detectados com dados de velocidade média, e verificaram que 81% dos eventos correspondem à realidade. Esta pesquisa forneceu esclarecimentos para o monitoramento das condições de tráfego em plataformas de transporte inteligente.

Com base nesses dados, é possível detectar incidentes no tráfego urbano e superar algumas limitações enfrentadas pelas tecnologias de transporte convencional. De modo geral, os planejadores de transportes podem se beneficiar com uso desses dados, uma vez que os sensores tradicionais apenas detectam mudanças nas medidas de tráfego, e são incapazes de identificar o tipo de evento que está ocorrendo, além de poder ter feedback dos usuários (JONES *et al.*, 2018).

#### 4.2.3.2 Mobilidade urbana e atividades

No planejamento de transportes, os dados das redes sociais podem ser empregados na identificação de padrões de mobilidade e atividades. A partir dos *check-ins* geolocalizados, esses fornecem informações a respeito das atividades realizadas ao longo do dia, dias da semana e meses. Desse modo, esses dados fornecem informações importantes para planejadores de transportes (NIKOLAIDOU e PAPAIOANNOU, 2018).

Chaniotakis e Antoniou (2015) exploraram o potencial dos *check-ins* do *Twitter* na área de transporte. O estudo foi realizado em Atenas, Grécia. Os resultados indicaram que os *check-ins* podem explicar o comportamento de viagens dos indivíduos, principalmente quando se trata de indivíduos de renda mais elevada. Além disso, foi percebido que há um maior agrupamento de viagens atraídas para áreas de lazer em horário não comercial.

Chaniotakis *et al.* (2016) extraíram *check-ins* do *Twitter*, *Foursquare* e *Facebook*, referentes à cidade de Thessalonki, Grécia. Foram empregados cinco meses de dados, referentes a aproximadamente 18.000 usuários. A metodologia aplicada consistiu em observar as atividades realizadas pelos usuários e suas características temporais. Os resultados evidenciaram que diferente da pesquisa convencional, os *check-ins* são geralmente efetuados em atividades do tipo lazer. Eles concluíram que esses dados podem servir de base para a área de transporte, no que se refere obter os destinos de viagens pelo motivo lazer.

Por outro lado, Abbasi *et al.* (2017) verificaram a aplicação de *check-ins* advindos do *Foursquare* na previsão de padrões de mobilidade em Manhattan, Nova York. Para isso, eles propuseram três cenários: distância de classificação, número de locais entre a origem e o destino e um esquema de ponderação de *check-ins* para calcular as classificações. A metodologia consistiu em um modelo conhecido como Rank, onde esse é baseado em classificação, sendo semelhante à estrutura do modelo de gravidade. Como esse modelo é parametrizado, foi necessário calibrá-lo para determinar o parâmetro de ajuste. Mas, para o método de calibração foi aplicado o modelo de *Hyman* na matriz, onde esse minimiza a diferença entre a distância real e a distância modelada de modo repetitiva. Após isso, o melhor parâmetro determinado foi empregado para calcular a distribuição do fluxo de mobilidade. Após isso, o autor comparou os resultados obtidos com os dados de viagens de táxi, sendo esses fornecidos pela Comissão de Táxis e

Limousimnes (TLC). Os resultados indicaram que os dois primeiros cenários resultaram em padrões semelhantes. Mas, o terceiro cenário apresentou uma melhor precisão das previsões. Portanto, os *check-ins* desempenham um papel significativo na melhoria da previsibilidade dos padrões de mobilidade, que pode servir de apoio aos planejadores de transporte.

Salas e Quezada (2017) validaram o uso do *Twitter* com a finalidade de mapear os padrões de mobilidade na cidade de Concepción, Chile. A metodologia se baseou no modelo de autocorrelação espacial bivariada. Os resultados evidenciaram uma quantidade de origens de viagens em áreas residenciais de alta renda situada ao norte do centro da cidade e na região sudoeste da cidade.

Pimpale *et al.* (2018) analisaram o uso do *Twitter* para identificar indivíduos que exerciam influência em tópicos relacionados a área do sistema de transporte. O estudo foi aplicado no condado da Califórnia, Estados Unidos. A metodologia consistiu no desenvolvimento de um site atrelado ao *Twitter*, onde cada usuário selecionava um tópico relacionado a área de transporte. Em seguida, os *tweets* eram recuperados em tempo real e relacionados aos indivíduos considerados influentes com relação ao tópico específico. O grau de influência foi avaliado a partir de duas medidas. A primeira sendo o número de menções no *Twitter* e a outra sendo a quantidade de *retweets*. Além disso, os *tweets* foram relacionados a um banco de dados do Google Maps. Os autores concluíram que essa metodologia pode ser utilizada por órgãos responsáveis pelo sistema de transportes. Além disso, pode encorajar a limitar a desinformação e aceitação e um novo serviço de transporte.

Carvalho *et al.* (2020) coletaram dados do *Twitter* para a compreensão dos padrões de mobilidade e atividades realizadas por eles na cidade do Porto, Portugal. Eles obtiveram aproximadamente 24876 *tweets*, do período de setembro de 2019 a fevereiro de 2020. As postagens foram divididas em diferentes grupos como: dias da semana, finais de semana, manhã, tarde e noite. Em seguida, identificaram os diferentes padrões de deslocamentos em termos espaciais e temporais. Os resultados apresentaram que existem pontos na cidade que são mais visitados do que outros locais. Esta análise, permitiu a compreensão da distribuição espacial das pessoas, além de descrever uma série de medidas relacionadas ao planejamento urbano. Uma das medidas seria tornar determinados locais da cidade mais atraentes, e assim, propor novas vias urbanas para o transporte público.



Miyazawa *et al.* (2020) avaliaram a relação entre informações textuais dos *tweets* e dados de GPS advindos de Smartphones no Japão. Eles empregaram um algoritmo para extrair informações textuais dos *tweets*. Em seguida, empregaram modelos de aprendizagem profunda com a finalidade de estimar diferentes padrões de mobilidade. Com base nos resultados, eles sugeriram que a combinação entre as bases de dados pode ser viável para a modelagem da dinâmica urbana.

Os trabalhos supracitados tiveram como foco principal a aplicação dos dados das redes sociais na demanda por transportes. Compreender essa finalidade pode fornecer aos pesquisadores, planejadores de transportes e órgãos governamentais, uma visão mais ampla de como se encontra o sistema de transporte em um dado local, além de fornecer ideias para novas implementações.

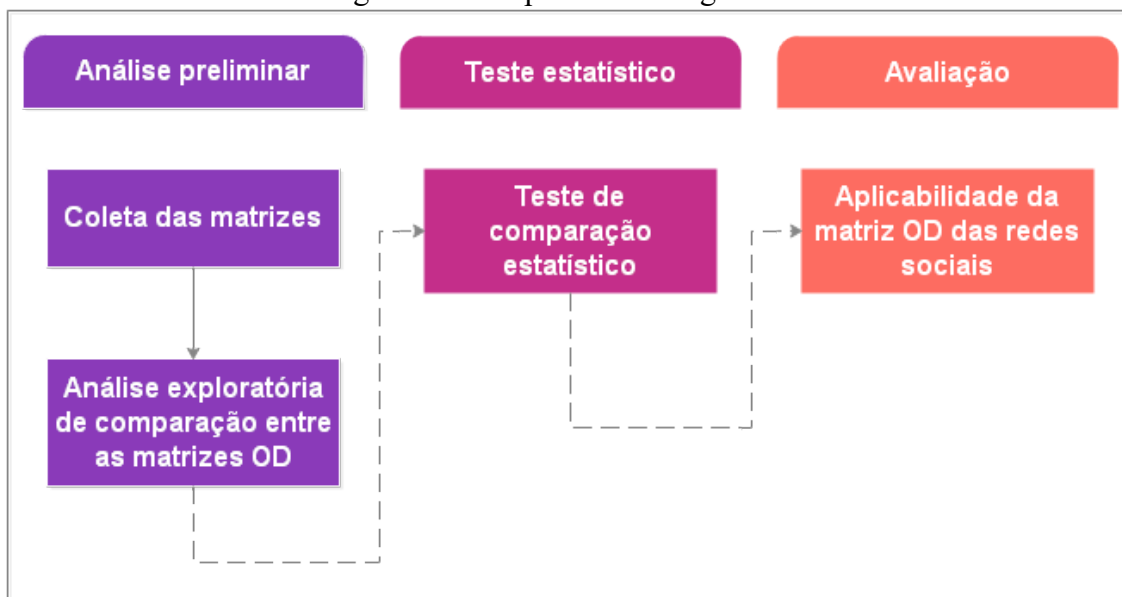
#### **4.2.4 Considerações**

Os trabalhos supracitados apontam inquietações sobre o potencial das redes sociais na obtenção de matrizes OD, assim como uma baixa representatividade dos dados e vieses socioeconômicos. Além disso, verifica-se ainda a falta de um direcionamento a respeito da aplicabilidade dessa matriz OD na prática. Entender esse propósito pode fornecer aos pesquisadores, informações sobre uma visão mais ampla do sistema de transporte em um determinado local e gerar ideias para novas implementações. Embora a literatura apresente essas restrições, alguns pesquisadores mostraram que estes dados podem sim ser considerados um subproduto valioso para os diversos estudos, e quando estes são georreferenciados, agregam valor à diversas investigações potenciais na área do planejamento de transportes.

### **4.3 Método**

Afim de se atingir o objetivo de comparação entre a matriz OD obtida a partir das redes sociais e a matriz OD modelada do TRANUS, realizou-se uma análise estatística, bem como uma avaliação da aplicabilidade da matriz OD advinda das redes sociais. Essa investigação considerou apenas dois motivos de viagens distintos: por motivo Trabalho e por motivo Outro. A Figura 21, ilustra a metodologia.

Figura 21 – Etapas Metodológicas



Fonte: Elaborado pelo autor.

#### 4.3.1 Obtenção das matrizes

Primeiramente, os dados que compuseram a matriz OD baseada nas redes sociais foram coletados tanto no *Instagram* quanto no *Twitter*. A matriz OD assim obtida, como descrito no Capítulo 2, considerou viagens por dois motivos, sendo eles: Trabalho e Outro, obviamente todos os motivos como educação e lazer, por exemplo, foram agrupados no motivo “Outro”. Durante recorte temporal usado para coleta dos dados das redes sociais, foram obtidos aproximadamente 1,37 milhões de *check-ins*, com base em 131.943 usuários das redes sociais. Após os processos de estruturação dos dados, foram obtidas 288.264 viagens, equivalentes a um total de 77215 usuários. As viagens assim obtidas, constituíram duas matrizes, a saber: Matriz OD por motivo Trabalho e Matriz OD por motivo Outro. Com base nessas viagens, houve a necessidade de se compreender quem seriam os usuários que realizaram tais viagens, e a partir de testes estatísticos espaciais, verificou-se que a maioria desses indivíduos pertence às classes de média e a alta renda.

Ressalte-se que devido à ausência de dados recentes coletados na área de estudo, através de uma pesquisa domiciliar e da impossibilidade de aguardo por uma nova pesquisa domiciliar, devido às restrições de tempo demandado para finalização desta investigação, foi necessário o uso, para efeito comparativo, de uma matriz OD externa

modelada. Tal matriz OD foi obtida a partir do uso do modelo integrado de transportes e uso do solo TRANUS, referente ao ano de 2015, para a mesma região estudada. Destaca-se ainda que esta matriz OD foi obtida a partir da calibração dos modelos de transportes do TRANUS para o ano base de 2000, com base em informações da pesquisa domiciliar de 1996. Em seguida os parâmetros iniciais foram empregados para calibração do ano base de 2015 considerando novas medidas de desempenho (SOUZA, *et al.*, 2019). A matriz OD do TRANUS foi modelada para a hora pico da manhã (entre 7:00 e 8:00) de um dado dia útil.

De forma análoga, a matriz OD das redes sociais obtida também corresponde a um dado dia útil. Vale ressaltar ainda que para todas as análises posteriores, considerou-se que os indivíduos se deslocaram através do modo motorizado individual uma vez que no Capítulo 2, as viagens obtidas corresponderam aos indivíduos pertencentes em sua grande maioria a média e a alta renda. Por outro lado, empregou-se duas matrizes OD do TRANUS, a saber: por motivo Trabalho da população de renda média e alta (modo motorizado) e uma outra matriz por motivo Outro (modo motorizado).

#### **4.3.2 *Análise exploratória da comparação entre as matrizes***

Essa etapa consistirá em uma análise de similaridade, ou seja, serão realizadas comparações entre as matrizes OD das redes sociais e as matrizes OD do TRANUS, levando em consideração o motivo da viagem. Inicialmente, as matrizes serão exemplificadas através de mapas de agrupamento de *Moran (LISA Cluster Map)*, elaborados a partir da plataforma computacional GeoDA. Assim, será possível identificar a distribuição espacial das origens e destino de ambas as matrizes. A partir dessas análises, ficará evidente as zonas em comum entre os mapas, e assim pode-se definir algum grau de similaridade entre elas.

#### **4.3.3 *Teste de estatístico de comparação***

Em seguida, serão aplicados dois testes estatísticos, sendo eles: o teste t-pareado e o de regressão linear. O primeiro terá como objetivo compreender se não há diferença entre as médias de viagens entre as duas matrizes OD em análise, com base no tipo de motivo. Para esse fim, as matrizes OD serão pareadas de forma alternada, levando

em consideração o motivo das viagens. A análise será realizada entre grupos de valores não independentes.

Sendo as hipóteses definidas como:

$$\begin{cases} H_o: \mu_{OTRDS} = \mu_{OTTNS}, \mu_{OORDS} = \mu_{OOTNS}, \mu_{DTRDS} = \mu_{DTTNS}, \mu_{DORDS} = \mu_{DOTNS} \\ H_a: \mu_{OTRDS} \neq \mu_{OTTNS}, \mu_{OORDS} \neq \mu_{OOTNS}, \mu_{DTRDS} \neq \mu_{DTTNS}, \mu_{DORDS} \neq \mu_{DOTNS} \end{cases}$$

A hipótese nula considera que não há diferença entre as médias populacionais, ou seja, a média de viagens entre as origens e destinos das matrizes são iguais. Desse modo, ao se comparar em dupla as origens de viagens da matriz OD das redes sociais (*OTRDS - Trabalho e OTTNS - Trabalho*) com as origens da matriz OD do TRANUS (*OORDS - Outro e OOTNS - Outro*), essas não irão variar significativamente segundo a hipótese nula. Do mesmo modo, ao se comparar os destinos de viagens da matriz das redes sociais (*DTRDS - Trabalho e DORDS - Outro*) com os destinos da matriz do TRANUS (*DTTNS - Trabalho e DOTNS - Outro*), essas tenderão a se comportar de modo semelhante, ou seja, não variar significativamente. Caso a hipótese nula seja rejeitada, existirá indícios para aceitação da hipótese alternativa. A hipótese alternativa indica que uma das matrizes se comportou de modo diferente.

Logo após a aplicação do teste t-pareado, será aplicado o teste de regressão linear. Esse tem como objetivo medir o quão semelhante as matrizes são, tendo o R-quadrado como base. No decorrer da análise, se o valor do R-quadrado for significativamente alto, isso indicará a existência de correlação entre as matrizes; caso contrário, indicará uma fraca correlação. Por fim, serão formuladas conclusões sobre os resultados dessa tentativa de comparação.

#### **4.3.4 Avaliação de aplicabilidade**

Com base nos resultados exploratórios e testes estatísticos, serão abordadas as conclusões das possíveis aplicabilidade dessa matriz OD na área do planejamento de transportes.

### **4.4 Resultados**

Os resultados aqui apresentados, foram baseados na matriz OD das redes sociais provenientes do Capítulo 2 e a matriz OD do TRANUS. A abordagem empregada permitiu identificar visual e estatisticamente as semelhanças existentes entre as matrizes OD das redes sociais e as matrizes OD do TRANUS, com base nas origens e destinos das viagens para ambos os motivos. Porém, as análises realizadas partiram da premissa de que as viagens obtidas a partir das redes sociais foram efetivadas pelo modo motorizado e em sua grande maioria por indivíduos pertencentes a média e alta renda.

#### **4.4.1 Análise exploratória**

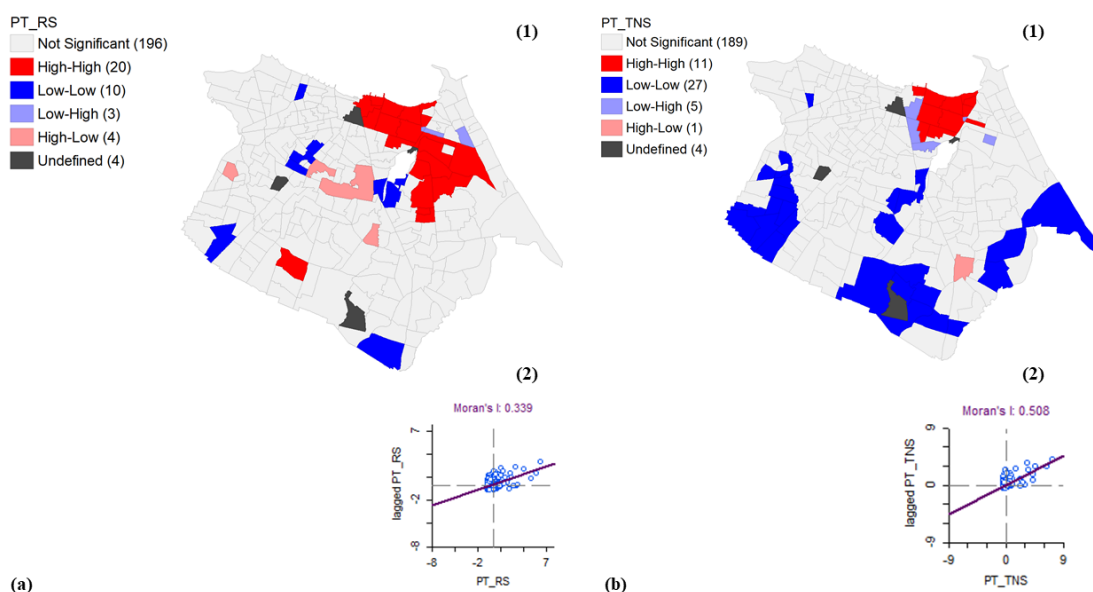
Inicialmente, foi realizada uma análise exploratória (visual) e comparativa entre as origens e destinos das matrizes OD das redes sociais com as origens e destinos das matrizes OD do TRANUS, com base no motivo da viagem e da renda. Assim, as matrizes OD do TRANUS empregadas nesta análise equivalem àquelas nas quais as viagens foram realizadas por indivíduos de média e alta renda com base no modo motorizado.

##### *4.4.1.1 Número de viagens por motivo Trabalho originadas em uma dada Zona de tráfego*

A Figura 22 apresenta o *LISA Cluster Map* (Figura 22a.1 e 22b.1) para as Origens por motivo Trabalho das duas matrizes. O *LISA Cluster Map* indica os agrupamentos estatísticos para a origem por motivo Trabalho, que foram obtidos a partir do critério de contiguidade (*queen*) para a matriz de proximidade (*W*), além do gráfico de dispersão com índice de *Moran* (Figura 22a.2 e 22b.2). A decisão de empregar essa matriz de proximidade (*queen*) se deu pela sua simplicidade de compreensão, visto que ela considera apenas as zonas que possuem uma fronteira em comum. As zonas de cor cinza representam as que não estão inclusas no estudo. Como se pode notar no mapa de agrupamento de *Moran* da Figura 22a.1, existe uma configuração espacial bem definida com um índice de autocorrelação de *Moran* de 0,34. Enquanto na Figura 22b.1, verifica-se um perfil espacial ainda mais significativo com um índice de *Moran* de 0,51. Ainda no mapa de agrupamento (Figura 22a.1), se observa um agrupamento de zonas em *High-High* (alto-alto, zonas com valores estatisticamente elevados da variável, vizinhas de zonas com valores igualmente altos) ao norte e a leste da cidade. Essas zonas são

habitadas predominantemente por indivíduos de renda mais elevada. As zonas *Low-Low* (baixo-baixo, zonas com valores estatisticamente baixos da variável vizinhas de zonas com valores igualmente baixos) estão espalhadas a oeste e sudoeste da cidade em regiões periféricas. Ao passo que, a Figura 22b.1 apresenta um agrupamento de zonas *High-High* predominante na área norte da cidade. Nesta região, encontram-se indivíduos pertencentes à classe de alta renda. As zonas *Low-Low* estão agrupadas a leste, oeste e sul da cidade. Esses agrupamentos ocorrem em áreas onde predominam em sua maioria indivíduos pertencentes à classe de baixa renda. De modo geral, existe um padrão semelhante entre os gráficos (Figuras) quando se refere às zonas *High-High*. Mas, quando se analisa as zonas *Low-Low*, verifica-se um padrão distinto. Isso significa que a Figura 22b.1 possui certos comportamentos que não são vistos na Figura 22a.1.

Figura 22 – Comparação entre *LISA Cluster Map* e gráfico de dispersão do modelo das origens de viagens das redes sociais (a) e o *LISA Cluster Map* e gráfico de dispersão do modelo das origens de viagens do TRANUS (b) por motivo Trabalho



Fonte: Elaborado pelo autor.

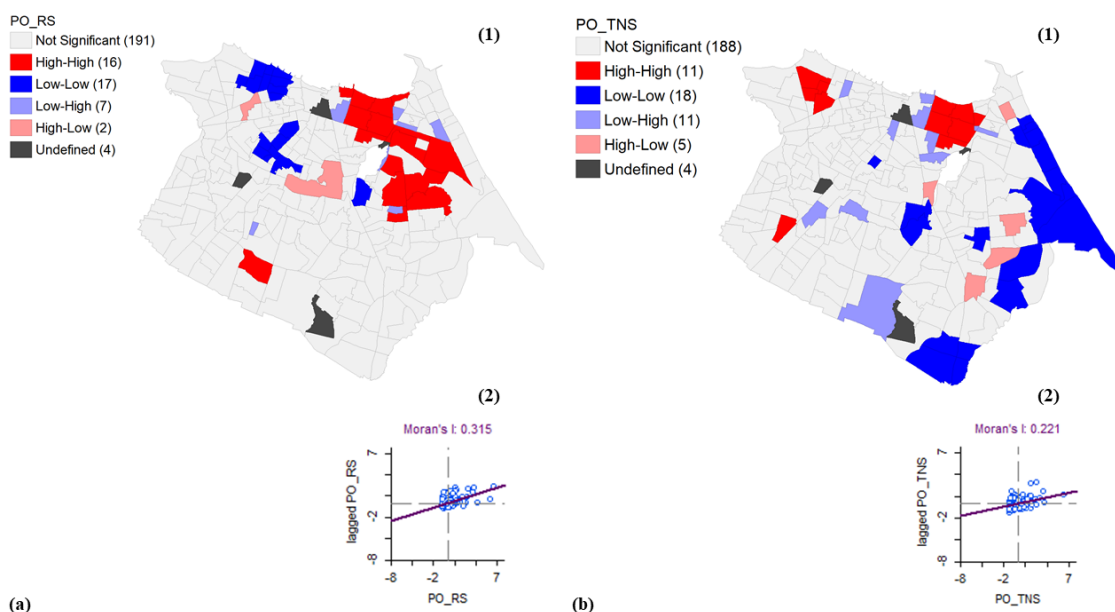
#### 4.4.1.2 Número de viagens por motivo Outro, originadas em uma dada Zona de tráfego

A partir da Figura 23a.1 verifica-se as zonas *High-High* a norte e a leste da cidade. Essas zonas e suas vizinhas realizam um número estatisticamente elevado de

viagens por motivo Trabalho, enquanto as zonas *High-Low* que estão localizadas a oeste são aquelas que realizam viagens por motivo Outro, e as suas zonas vizinhas possuem efeito contrário. Essas zonas são adensadas por domicílios de média e alta renda. Além disso, se pode notar que algumas zonas a oeste da cidade aparecem em *Low-Low*. Essas detêm características contrárias às zonas *High-High*, apresentando uma tendência a não realizar viagens por esse motivo. Por outro lado, a partir Figura 23b.1 observa-se as zonas *High-High* a norte, leste e oeste exclusivamente. Enquanto as zonas *High-Low* se agrupam a sudeste da cidade, diferente da Figura 23a.1. Do mesmo modo, as zonas *Low-Low* se comportam de forma diferente da Figura 23a.1, ou seja, elas se agrupam a leste e a sudeste. Em seguida, obteve-se o índice de *Moran* de 0,31 (Figura 23a.2), o que indica a existência de correlação espacial. Em contrapartida, na Figura 23b.2 observa-se um índice de *Moran* de aproximadamente 0,22.

De modo geral, as Figuras possuem alguns padrões semelhantes. Mas, a Figura 23a.1 possui algumas zonas *High-High* a mais do que a Figura 23b.1. Isso indica que as zonas que se encontram a leste, originam viagens pelo motivo Outro, e que não é possível identificar ao se analisar a Figura 23b.1.

Figura 23 - Comparação entre *LISA Cluster Map* e gráfico de dispersão do modelo das origens de viagens das redes sociais (a) e o *LISA Cluster Map* e gráfico de dispersão do modelo das origens de viagens do TRANUS (b) por motivo Outro



Fonte: Elaborado pelo autor.

#### 4.4.1.3 Número de viagens por motivo Trabalho destinadas a uma dada zona de tráfego

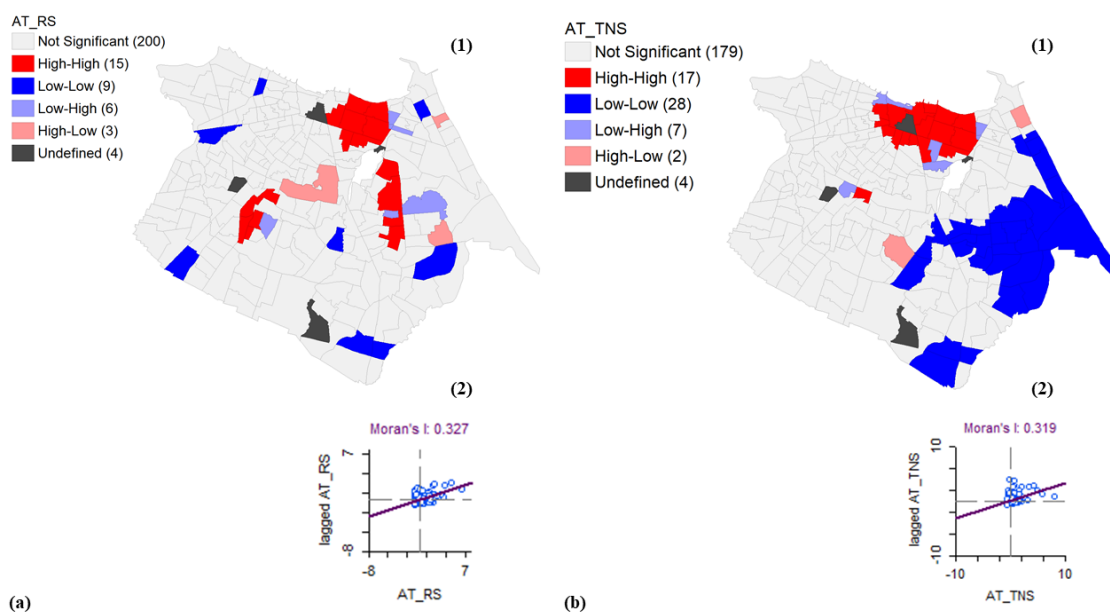
Nas Figuras 24a.1 e 24b.1, se apresentam o *LISA Cluster Map*, mapeando os agrupamentos espaciais das viagens atraídas por motivo Trabalho. Como se pode observar na Figura 24a.1, ao norte, sudeste e uma zona a sudoeste da cidade, há agrupamentos do tipo *High-High* da variável número de viagens atraídos. Isso significa que essas são regiões que mais atraem viagens na cidade. Enquanto a Figura 24b.1 apresentam zonas do tipo *High-High* ao norte e centro exclusivamente. O padrão em ambas as Figuras são plausíveis, uma vez que esses agrupamentos abrangem o centro da cidade, áreas comerciais importantes e *shoppings centers*. Essas regiões concentram grande parte dos empregos da cidade.

Por outro lado, na Figura 24a.1 verifica-se que as zonas em *Low-Low* se encontram espalhadas exclusivamente em áreas periféricas por toda a cidade. Enquanto na Figura 24b.1, essas zonas encontram-se a leste, sudeste e sul da cidade. Diante do nítido padrão de agrupamentos concentrados em regiões específicas, o índice de *Moran* da Figura 24a.2 foi equivalente 0,33, enquanto na Figura 24b.2 foi 0,32. Esses valores indicam uma significativa autocorrelação espacial da variável o que de fato pode ser observado através do *LISA Cluster Map* e do mapa de significância de *Moran*.

As Figuras 24a.1 e 24b.1 possuem uma grande distinção quando se analisa as zonas *Low-Low*. Pois, a Figura 24a.1 possui um padrão que se distribui ao longo do território, enquanto a Figura 24b.1 apresenta um agrupamento exclusivo a leste.



Figura 24 - Comparação entre *LISA Cluster Map* e gráfico de dispersão do modelo das destino de viagens das redes sociais (a) e o *LISA Cluster Map* e gráfico de dispersão do modelo dos destinos de viagens do TRANUS (b) por motivo Trabalho



Fonte: Elaborado pelo autor.

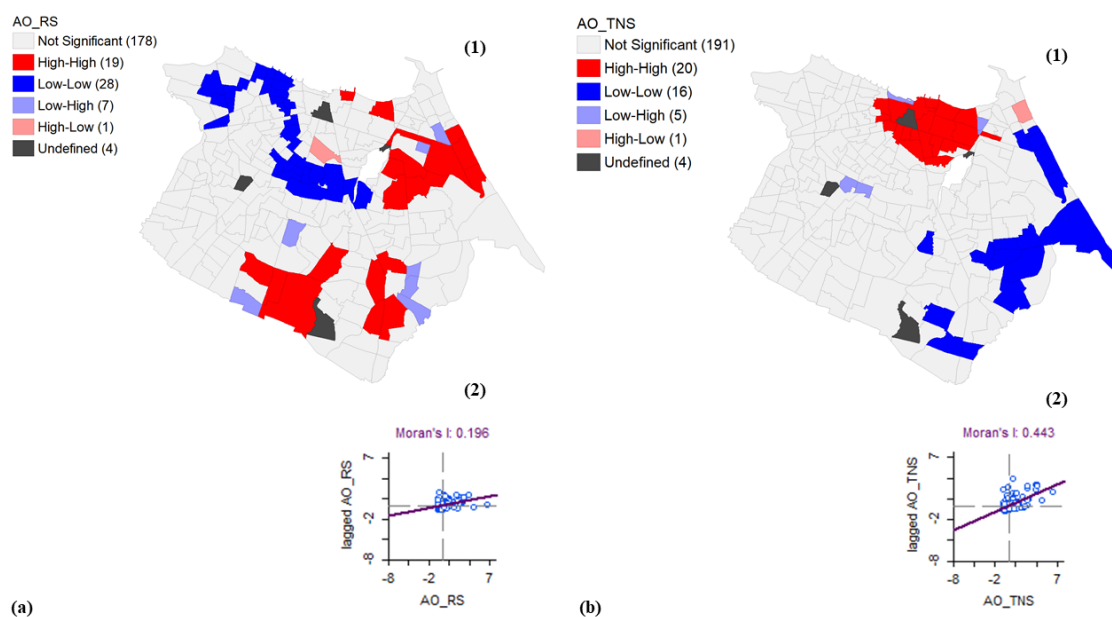
#### 4.4.1.4 Número de viagens por motivo Outro, destinadas a uma dada zona de tráfego.

A caracterização do número de viagens atraídas por motivo Outro consistiu em uma análise exploratória da variável, como explicado na apresentação do método. Essa análise foi feita com base na Figura 25, que apresenta o *LISA Cluster Map* (Figura 25a.1). A partir da Figura 25a.1, se verifica que as zonas do tipo *High-High* estão localizadas em boa parte da região norte, sudeste e sudoeste da cidade. Enquanto na Figura 25b.1, as zonas do tipo *High-High* se agrupam ao norte e centro da cidade. De modo geral, existe uma baixa relação entre as duas matrizes quando se refere a destino de viagens por motivo Outro. Nessas zonas *High-High*, encontram-se diversos locais onde as pessoas costumam ir para realizar outras atividades que não sejam trabalho, como fazer compras, ou seja, nelas se concentram diversos locais de lazer, cuidado com a saúde, centro de eventos e de convenções, supermercados, *shopping's centers*, farmácias, restaurantes e academias. Ainda com base na Figura 25a.1, existem outras zonas em *High-High* espalhadas na região sul da cidade que concentram alguns estabelecimentos públicos importantes, como o hospital Sarah Kubitscheck e o estádio Castelão. Por outro

lado, as zonas em *Low-Low* estão espalhadas em pequenas zonas residenciais na região noroeste da cidade.

De modo geral, essas duas Figuras mostraram-se diferentes. Talvez os destinos das viagens com base nos dados de redes sociais não possam ser explicados a partir dos destinos da matriz OD do TRANUS.

Figura 25 - Comparação entre *LISA Cluster Map* e gráfico de dispersão do modelo dos destinos de viagens das redes sociais (a) e o *LISA Cluster Map* e gráfico de dispersão do modelo dos destinos de viagens do TRANUS (b) por motivo Outro



Fonte: Elaborado pelo autor.

Com base nas análises, pode-se verificar que os padrões entre origens e destinos das matrizes se assemelham, porém com exceção dos destinos por motivo Outro. Este último apresentou um comportamento diferente dos demais. O destino por motivo Outro difere ao mostrar que zonas de destino não se assemelham às da matriz OD do TRANUS, porém algumas zonas *Low-Low* aparece em uma e na outra não. Isso sugere que os usuários das redes sociais são um pouco diferentes dos usuários modelados no TRANUS. De modo geral, apesar de existir uma ligeira semelhança entre as matrizes, ainda não está claro o grau de semelhança entre elas. Assim, o próximo passo será a aplicação de testes estatísticos para comparar e concluir se existe ou não semelhança entre elas.

#### 4.4.2 *Análise comparativa*

Neste tópico são apresentados os resultados acerca do teste t-pareado e do teste de regressão linear, esse por último, teve como finalidade medir o nível de relação entre os pares.

Os resultados obtidos a partir da comparação entre os pares são apresentados nas Tabelas 6, 7, 8 e 9. A Tabela 6 apresenta uma comparação entre o número de viagens por motivo Trabalho originadas em uma dada zona. Nesse caso foram utilizadas as origens das viagens pelo motivo Trabalho da matriz OD das redes sociais (*OT\_RDS*) e da matriz OD do TRANUS (*OT\_TNS*). Já a Tabela 7 exibe a comparação entre o número de viagens, por motivo Outro, originadas em uma dada zona. Para essa comparação, foram empregadas as origens das viagens pelo motivo Outro da matriz OD das redes sociais (*OO\_RDS*) e da Matriz OD do TRANUS (*OO\_TNS*). A Tabela 8 traz uma comparação entre o número de viagens por motivo Trabalho destinadas a uma dada zona de tráfego. A análise comparativa foi realizada entre os destinos de viagens pelo motivo Trabalho da matriz OD das redes sociais (*DT\_RDS*) e da matriz OD do TRANUS (*DT\_TNS*). Por fim, a Tabela 9 apresenta uma comparação entre os destinos de viagens pelo motivo Outro da matriz OD das redes sociais (*DO\_RDS*) e da matriz OD do TRANUS (*DO\_TNS*).

Com base nos resultados das Tabelas, verificou-se que apenas a Tabela 9 apresentou um valor da estatística t acima do t crítico bi-caudal de 1,97, com um coeficiente de determinação  $R^2=0,09$ , onde esse valor é considerado baixo. Isso significa que os destinos de viagens por motivo Outro das redes sociais não possuem relação com os destinos por motivo Outro do TRANUS. Os demais resultados como apresentam as Tabelas 6, 7 e 8 foram significativos. A relação entre as origens por motivo trabalho da matriz OD das redes sociais e a matriz OD motivo trabalho do TRANUS foi de aproximadamente 0,52. Enquanto, a relação entre as origens da matriz OD das redes sociais por motivo Outro e a matriz OD do TRANUS foi de 0,42. Por fim, os destinos de viagens por motivo trabalho entre as matrizes OD foi de 0,33.

A divergência verificada entre a matriz OD da redes sociais por motivo Outro e a matriz do TRANUS por motivo Outro, pode estar relacionada ao fato de que o “Outro” é um motivo que agrega diversas atividades e isto pode ter concorrido para que os resultados fossem discrepantes, ou seja, o que está dentro do grupo Outro em uma matriz

e em outra matriz podem e certamente não sejam exatamente os mesmos motivos e por conseguinte as mesmas classes de usuários.

Tabela 6 – Teste t

	<b>OT_RDS</b>	<b>OT_TNS</b>
Média	11,56	10,06
Variância	251,4	483
gl	232	
Stat t	1,32	
P(T<=t) bi-caudal	0,1873	
t crítico bi-caudal	1,97	

Fonte: Elaborado pelo autor.

Tabela 7 – Teste t

	<b>OO_RDS</b>	<b>OO_TNS</b>
Média	10,09	9,06
Variância	197	386
gl	232	
Stat t	1,73	
P(T<=t) bi-caudal	0,13	
t crítico bi-caudal	1,97	

Fonte: Elaborado pelo autor.

Tabela 8 – Teste t

	<b>DT_RDS</b>	<b>DT_TNS</b>
Média	13,4	12,2
Variância	203	432
gl	232	
Stat t	1,67	
P(T<=t) bi-caudal	0,15	
t crítico bi-caudal	1,97	

Fonte: Elaborado pelo autor.

Tabela 9 – Teste t

	<b>DO_RDS</b>	<b>DO_TNS</b>
Média	8,74	5,80
Variância	144	51
gl	233	
Stat t	3,41	
P(T<=t) bi-caudal	0,0007	
t crítico bi-caudal	1,97	

Fonte: Elaborado pelo autor.

No entanto, é importante ressaltar que essa tentativa de comparação entre as matrizes não foi feita utilizando dados oficiais de pesquisas origem e destino, mas sim com os resultados de um outro esforço de modelagem e que isto só foi feito, devido à

indisponibilidade de pesquisas domiciliares de origem e destino em Fortaleza desde 1996 e também devido ao fato de este trabalho utilizar dados que surgiram recentemente. Porém, sabe-se que a cidade de Fortaleza passou por uma pesquisa domiciliar em 2019, e o fato de que até dezembro de 2020 ainda não foram divulgados seus resultados, ratifica ainda mais uma das desvantagens desse método, relacionada ao tempo de coleta e análise dos dados. Motivo pelo qual foi necessário usar os dados oriundos da modelagem integrada de transportes e uso do solo, o TRANUS.

Ainda que a matriz do TRANUS tenha sido obtida através de um esforço complexo de modelagem, não deixa de ser um modelo e nesse sentido, apenas uma tentativa de representar a realidade. Portanto, com nos resultados obtidos, pode-se afirmar que os dados de redes sociais podem ser empregados para as diversos estudo no planejamento de transportes, como exemplo, o estudo dos padrões de mobilidade na região estudada. Contudo, com esses resultados, levantam-se outras questões sobre, por exemplo, quem seriam esses usuários de redes sociais e quais grupos sociais eles representam, uma vez que se acreditava ser possível capturar o comportamento dos grupos de renda mais elevada da cidade a partir desses dados. Talvez esse seja um subgrupo ainda mais específico de pessoas de renda mais elevada, de uma determinada faixa etária e que vive um estilo de vida mais específico.

#### ***4.4.3 Avaliação de aplicabilidade das Matrizes OD de Redes Sociais***

Com base nos resultados exploratórios e estatísticos, verificou-se que as matrizes OD das redes sociais possuem uma relação positiva com as matrizes OD do TRANUS, com exceção dos destinos de viagens, por motivo Outro, que apresentou uma baixa correlação. Porém, com base nos resultados, verifica-se a necessidade de avaliar as possíveis aplicabilidades dessa matriz OD das redes sociais.

De modo geral, tais matrizes OD podem ser empregadas, por exemplo, para tentar compreender as dinâmicas de viagens urbanas em diferentes escalas temporais, como também, detectar congestionamentos de tráfego urbano e fluxos de deslocamento. Na área do planejamento de transportes, a matriz OD é uma ferramenta fundamental para a realização de diagnóstico, prevendo a demanda de viagens e contribuindo para a modelagem de transporte e otimização da rede. Os fluxos entre origem e destino, fornecem informações úteis para a previsão de serviços e estratégias de controle, como

por exemplo, para fins de modelagem. Mas, além das origens e destinos, é necessário identificar as preferências modais (RASHIDI *et al.*, 2017). Porém, sabe-se que a partir das matrizes OD das redes sociais, a escolha modal é uma de suas limitações (RASHIDI *et al.*, 2017).

Tradicionalmente, a matriz OD é obtida através de pesquisas domiciliares, contagens volumétricas, etc. No entanto, são consideradas caras, demoradas e estáticas. Mas, como apresentado nesta investigação, os dados oriundos de redes sociais, do *Twitter* e do *Instagram* em particular, fornecem uma nova maneira de estimar uma matriz OD, que pode servir de base complementar à pesquisa domiciliar.

Arjona e Palomares (2019) obtiveram uma matriz OD a partir dos dados do *Twitter* e identificaram os destinos pelo motivo Trabalho. O estudo foi realizado na área metropolitana de Madri como estudo de caso. Assim como eles, nessa Tese foi estimado os destinos de viagens pelo motivo Trabalho, levando em consideração o uso do solo e o horário do *check-ins*. Além disso, sabe-se que ao realizar a comparação entre a matriz OD das redes sociais e a matriz OD do TRANUS, essa mostrou-se ter um comportamento parecido. Segundo Pajevic e Shearmur (2017), obter uma matriz OD pelo motivo trabalho, não é apenas de interesse acadêmico, mas de interesse prático por urbanistas e planejadores de transportes. Ela fornece informações de como os trabalhadores se movem ao longo do dia, além disso, pode servir de *input* para o gerenciamento de sistemas de transportes e de espaços públicos.

Por outro lado, segundo a literatura a matriz OD das redes sociais pode ser empregada na etapa de distribuição de viagens, sendo essa uma das etapas do modelo sequencial, ou melhor, do modelo de quatro etapas. A etapa de distribuição de viagens busca alocar as viagens produzidas e atraídas, e esse processo proporciona uma melhor compreensão dos fluxos de viagens dentro do meio urbano (FAGNANT e KOCKELMAN, 2014). Pourebrahim *et al.* (2018) investigaram a aplicabilidade de uma matriz OD a partir de dados do *Twitter* na previsão da distribuição de viagens na cidade de New York, EUA. Eles aplicaram os fluxos de viagens em dois modelos, redes neurais e o gravitacional. Os resultados apresentaram que a matriz OD fruto do *Twitter* pode servir na modelagem de transporte, mais precisamente na previsão de viagens. Os autores verificaram que ao aplicar os fluxos de viagens em ambos os modelos, pode-se observar uma diminuição do erro quadrado e um aumento do  $R^2$ . Além disso, verificou-se que o modelo gravitacional superou o de redes neurais. Com base neste trabalho, sabe-se que as matrizes das redes sociais como o *Instagram* e o *Twitter* podem ser aplicadas com a

finalidade de compreender os diferentes fluxos dentro do meio urbano, servindo de base para o planejamento de transportes. Além desse tipo de aplicação, a matriz OD das redes sociais pode ser empregada na previsão de tráfego (Beiro *et al.*, 2016), Padrões de mobilidade (Kurkcu *et al.*, 2016) e fluxos migratórios (Hsiao *et al.*, 2020). Apesar dessas vantagens, deve-se ressaltar que esse tipo de matriz possui diversos vieses, socioeconômicos, sexo, idade, etc. Esses vieses precisam ser considerados e mencionados ao lidar com esse tipo de fonte de dados (CEBELAK, 2015). Em contrapartida, esse tipo de matriz OD possui como vantagem poder servir de input para diversos estudos em transportes, em especial de servir de complemento a matriz OD domiciliar.

#### 4.5 Considerações finais

Neste capítulo, foi realizada uma comparação entre a matriz OD desenvolvida a partir das redes sociais e a matriz OD modelada do TRANUS. Inicialmente, as matrizes foram coletadas, sendo elas, a matriz OD das redes sociais obtida no Capítulo 2 e uma matriz externa que foi obtida através de uma modelagem integrada de transportes e uso do solo, o TRANUS. Em seguida, foi realizada uma análise de similaridade prévia entre as matrizes com uso de mapas de agrupamento de *Moran (LISA Cluster Map)*. Após essa análise prévia, foram aplicados dois testes estatísticos, o teste t-pareado e o teste de regressão linear. Baseado nessas análises, foi possível obter resultados finais acerca da semelhança entre as matrizes. Por fim, foram abordados possíveis aplicabilidades dessa matriz OD no planejamento de transportes.

Os resultados prévios mostraram que existe uma similaridade entre as matrizes OD, com exceção dos destinos por motivo Outro das redes sociais, esses mostraram-se diferentes. Uma das possíveis explicações para essa discrepância é a de que as atividades, por motivo Outro, realizadas pelos usuários das redes sociais sejam diferentes das atividades dos usuários consideradas pelo TRANUS. Em seguida, os testes estatísticos, tais como o teste t-pareado e o teste de regressão identificaram que as matrizes possuem similaridade e correlação, apenas os destinos por motivo Outros diferiram, com um  $R^2$  de 0,09. Com base nos demais resultados, verifica-se as diferentes possibilidades de aplicações das matrizes OD na engenharia de transportes. A partir da matriz OD das redes sociais pelo motivo Trabalho, essa pode fornecer informações de como os trabalhadores se deslocam ao longo do dia, semanas e meses. Esse tipo de informação pode servir de input para o gerenciamento dos sistemas de transportes. Por

outro lado, essa matriz OD pode ser empregada de entrada na distribuição de viagens, e assim, ser possível compreender melhor os fluxos de viagens dentro do meio urbano. Embora a matriz OD das redes sociais possa ser utilizada em diversas aplicações na área do planejamento urbano, deve-se destacar que esse tipo de matriz possui alguns vieses, sendo eles socioeconômicos, de gênero, idade, etc. Ao lidar com essas fontes de dados passivos, esses vieses precisam ser considerados e mencionados. Um outro ponto a ser mencionado é a de que a matriz OD externa empregada foi realizada a partir de uma estimativa para o ano base de 2000, com informações da pesquisa domiciliar de 1996. Onde esses parâmetros iniciais foram utilizados para calibrar o ano de 2015. E os dados das redes sociais são referentes ao ano de 2018. Assim, verifica-se a existência de uma defasagem temporal entre as duas matrizes, além disso, a matriz do TRANUS foi originada a partir de estimativas e premissas diferentes. Apesar do resultado da comparação ter sido satisfatório, faz-se necessário realizar uma comparação a partir de dados advindos da pesquisa domiciliar de 2019, que até o momento não foi divulgada.



## 5 CONCLUSÕES

Tradicionalmente, o planejamento de transportes emprega a pesquisa OD domiciliar para a estimação da matriz OD. Tal abordagem apresenta algumas limitações relativas tanto à operacionalização da pesquisa quanto à qualidade das informações obtidas. Além disso, apresenta algumas desvantagens, tais como elevado custo financeiro e de tempo (normalmente mais de um ano). Ademais, esse tipo de pesquisa carece de constantes atualizações, sendo estas realizadas em uma frequência de 5 a 10 anos. Como exemplo, a cidade de Fortaleza não tem uma pesquisa do gênero há aproximadamente 24 anos. Tais limitações podem ser mitigadas a partir do uso de novas fontes de dados como, por exemplo, as redes sociais. Estas possibilitam a geolocalização dos seus usuários no momento em que esses compartilham os seus *check-ins*. Esse tipo de dado apresenta a grande vantagem de ser coletado de imediato e, quando combinado com dados socioeconômicos e de uso do solo, desponta como potencial fonte de informação para estimar padrões de mobilidade no meio urbano.

A partir dessa contextualização, esta pesquisa teve como objetivo geral analisar a aplicabilidade dos dados provenientes de redes sociais como o *Twitter* e o *Instagram* para a obtenção de uma matriz OD na cidade Fortaleza-CE. Objetivo esse que foi atingido com sucesso nesta investigação, na qual foram respondidas as questões de pesquisas que a embasam, e são apresentadas no tópico a seguir, bem como as suas principais conclusões.

### 5.1 Questões de pesquisa e conclusões

Diante das lacunas encontradas na literatura com relação a um método capaz de obter uma matriz OD, compreender o grau de variação das viagens entre dias úteis e as possíveis aplicabilidades da matriz OD, surgiram as questões de pesquisa. Essas estão elencadas e respondidas baseadas nos Capítulos 2, 3 e 4.

- *Como definir e caracterizar os deslocamentos a partir de dados geolocalizados advindos do Instagram e do Twitter?*

Com base nos dados do *Instagram* e do *Twitter*, foi possível definir e caracterizar os deslocamentos entre origens e destinos. Essas informações são valiosas para etapas posteriores, como a modelagem da demanda por viagens. Então, de posse dos dados e a partir de critérios definidos, foram inferidos as origens (domicílios) e os destinos. Variáveis de uso do solo serviram de apoio para inferir os motivos dos destinos e, a partir dessas origens e com o uso de variáveis socioeconômicas, foi identificado o perfil socioeconômico dos usuários.

Os resultados revelaram que o método foi eficiente em identificar as origens (domicílios) e os destinos de viagens, ou seja, em estimar a matriz OD. Além disso, foram inferidos os motivos de viagens, sendo eles: Trabalho e Outro. Análises estatísticas apresentaram que os domicílios amostrados estão localizados em regiões de média e alta renda. Desse modo, esses dados podem não representar viagens efetuadas por usuários de renda mais baixa em Fortaleza. Uma explicação para isso é a de que pessoas com esse perfil, em sua maioria, não têm acesso à internet de redes móveis e nem ao smartphone. Em tese, esses dados se mostraram promissores para inferir a mobilidade dos usuários de média e alta renda na cidade de Fortaleza.

- *Como as viagens baseadas em redes sociais variam ao longo dos dias e finais de semana?*

A partir dos dados do *Instagram* e do *Twitter*, foi possível entender o comportamento de viagens, ou melhor, a variabilidade de viagens na cidade de Fortaleza. Compreender a variabilidade pode interferir positivamente em melhores esquemas de gerenciamento de demanda na cidade. Portanto, uma metodologia foi desenvolvida para analisar o grau de variação do comportamento de viagens entre dias úteis e finais de semana, ao longo de oito meses. Hipóteses foram levantadas a partir de uma revisão da literatura e investigadas com base em testes estatísticos como o ANOVA e o Tukey.

Os resultados indicaram que existe uma transição entre os dias úteis e os finais de semana, mais precisamente nos dias equivalentes às quintas-feiras e sextas-

feiras. Como critério que atesta a confiabilidade desses dados, pode-se verificar uma maior frequência de viagens por motivo Trabalho durante a semana, e o oposto nos finais de semana. Ao comparar a média de viagens, por motivo Outro, entre sábados e domingos constatou-se não haver variações significativas. Além disso, verificou-se que a sexta-feira possui uma tendência a ter uma média maior de viagens e que a quantidade de viagens dos dias da semana difere do domingo para ambos os motivos. Por fim, observou-se que a quantidade de viagens do domingo difere do sábado e uma razão para isso é que os tipos de atividades realizadas em ambos os dias são diferentes. De uma perspectiva fenomenológica, verificou-se que as viagens baseadas em redes sociais podem explicar a variabilidade de viagens entre dias. Porém, é preciso levar em consideração as características de quem são esses usuários, explorado no Capítulo 2.

- *Como investigar a semelhança entre a matriz OD das redes sociais e a matriz OD externa?*

Com base em testes estatísticos foi possível comparar as matrizes. Primeiramente, as matrizes foram coletadas e em seguida foi observado à similaridade entre elas, representadas com base nos mapas de agrupamento de *Moran (LISA Cluster Map)*. De acordo com a análise anterior, foram aplicados dois testes estatísticos, e verificado as possíveis similaridades entre as matrizes OD. Por fim, com base na matriz OD das redes sociais, foi discutida as possíveis aplicabilidades dessa matriz. Os resultados dos testes estatísticos deixam clara a existência de similaridade entre as matrizes, com exceção dos destinos de viagens, por motivo Outro, que foram diferentes, com um  $R^2$  de 0,09. Com base nos resultados, é possível analisar possíveis aplicabilidades da matriz OD das redes sociais na engenharia de transportes. Um exemplo a ser mencionado é a matriz OD por motivo Trabalho, essa pode fornecer informações acerca dos deslocamentos dos trabalhadores dentro do meio urbano, servindo como input para o gerenciamento de sistemas de transportes. Além disso, essa pode servir de entrada para o modelo de distribuição, e assim, ser possível compreender os fluxos de viagens dentro do meio urbano. Embora essa matriz possa ser empregada em diversos campos do planejamento urbano, deve-se destacar os seus vieses e limitações.

## 5.2 Sugestões para trabalhos futuros

Para trabalhos futuros de pesquisa, seguem algumas sugestões:

- Buscar novas fontes de dados, tais como outras redes sociais, além do *Twitter* e do *Instagram*, e dados de bilhetagem eletrônica, na tentativa de incluir usuários do sistema de transporte urbano de todas as classes sociais;
- Insistir e testar outras formas de modelagem desagregada;
- Buscar novas variáveis explicativas para a etapa da geração de viagens;
- Aplicar recursos de inteligência artificial para que se possa realizar a leitura dos textos e inferir o tipo de atividade, o que tornaria dispensável a preexistência de mapeamento em nível de lote de dados acerca do uso do solo;
- Modelar a geração de viagens de forma desagregada;
- Aplicar um método de expansão a matriz OD de redes sociais e buscar validá-la com base em um estudo externo com informações atualizadas de pesquisas domiciliares e com recorte temporal homogêneo.

## 5.3 Publicações no período

OLIVEIRA, S. F. C.; SILVA, C. A. U.; OLIVEIRA, S. M. Y. **O MAPEAMENTO DO USO DO SOLO ATRAVÉS DO *TWITTER* COMO APOIO AO PLANEJAMENTO DE TRANSPORTES, O CASO DA CIDADE DE FORTALEZA-CE.** In: XXXI ANPET Congresso Nacional de Pesquisa e Ensino em Transportes, 2017, Recife. Anais do XXXI ANPET Congresso Nacional de Pesquisa e Ensino em Transportes, 2017.

FILHO, F. E. M.; CASSIANO, D. R.; OLIVEIRA, S. F. C.; BARBOSA, J. B. **ESTRUTURA URBANA E ESCOLHA MODAL: ESTUDO DE CASO DA UFC - CAMPUS DO PICI.** Revista dos Transportes Públicos, 2017.

OLIVEIRA, S. F. C.; SILVA, C. A. U.; OLIVEIRA, S. M. Y. **O USO DO *INSTAGRAM* E DO *TWITTER* PARA IDENTIFICAÇÃO E MAPEAMENTO DE USOS DO SOLO EM ÁREAS URBANAS: O CASO DE FORTALEZA-CE/THE USE OF *INSTAGRAM* AND *TWITTER* FOR IDENTIFICATION AND MAPPING OF SOIL USES IN URBAN AREAS: THE CASE OF FORTALEZA-CE.** Brazilian Journal of Development, 5(9), 16128-16149, 2019.

**OLIVEIRA, S. F. C.; SILVA, C. A. U. CARACTERIZAÇÃO DE PADRÕES DE DESLOCAMENTOS URBANOS EM FORTALEZA COM A UTILIZAÇÃO DE DADOS DE REDES SOCIAIS GEORREFERENCIADAS.**  
Revista Transportes (Rio de Janeiro), v. 28, p. 34, 2020.

## REFERÊNCIAS

ABBASI, O. R.; ALESHEIKH, A. A.; SHARIF, M. Ranking the city: the role of location-based social media check-ins in collective human mobility prediction. **ISPRS International Journal of Geo-Information**, v. 6, n. 5, p. 136, 2017.

AGRYZKOV, T.; MARTÍ, T.; TORTOSA, L. Measuring urban activities using Foursquare data and network analysis: a case study of Murcia (Spain). **International Journal of Geographical Information Science**, v. 31, n. 1, p. 100-121, 2017.

AGRYZKOV, T.; NOLASCO-CIRUGEDA, A.; OLIVER, J. L.; SERRANO-ESTRADA, L.; TORTOSA, L.; VICENT, J. F. Using data from Foursquare Web Service to represent the commercial activity of a city. *International Journal of Computer, Control, Quantum and Information Engineering*. **World Academy of Science, Engineering and Technology**, v. 9, n.1, p. 69-76, 2015.

AHAS, R.; AASA, A.; SILM, S.; TIRU, M. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. **Transportation Research Part C: Emerging Technologies**, v. 18, n. 1, p. 45-54, 2010.

AHMOUDA, A; HOCHMAIR, H. H.; CVETOJEVIC, S. Using Twitter to Analyze the Effect of Hurricanes on Human Mobility Patterns. **Urban Science**, v. 3, n. 3, p. 87, 2019.

ALLSTROM, A.; KRISTOFFERSSON, I.; SUSILO, Y. Smartphone based travel diary collection: Experiences from a field trial in Stockholm. **Transportation research procedia**, v. 26, p. 32-38, 2017.

ALZHRANI, S. M. **Development of IoT mining machine for Twitter sentiment analysis: mining in the cloud and results on the mirror**. Learning and Technology Conference, IEEE, p. 86-95, 2018.

ANSELIN, L. Local indicators of spatial association - LISA. **Geographical analysis**, 27, pp. 93-115, 1995.

ARJONA, J. O.; PALOMARES, J. C. G. Social media and urban mobility: Using Twitter to calculate home-work travel matrices. **Cities**, 89, p. 268-280, 2019.

ATROSTIC, B. K.; BATES, N.; BURT, G.; SILBERSTEIN, A.; WINTERS, F. **Nonresponse in federal household surveys: new measures and new insights**. In International Conference on Survey Nonresponse, 1999.

BARTOSIEWICZ, B.; PIELESIAK, I. Spatial patterns of travel behaviour in Poland. **Travel Behaviour and Society**, 15, 113-122, 2019.

BEIRO, M. G.; PANISSON, A.; TIZZONI, M.; CATTUTO, C. **Predicting human mobility through the assimilation of social media traces into mobility models**. EPJ Data Science, v. 5, n. 1, p. 30, 2016.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal statistical society: series B (Methodological)**, v. 57, n. 1, p. 289-300, 1995.

BIFET, A.; FRANK, E. **Sentiment knowledge discovery in Twitter streaming data**. In International conference on discovery science, pp. 1-15, 2010.

BISANZIO, D.; KRAEMER, M. U.; BREWER, T.; BROWNSTEIN, J. S.; REITHINGER, R. Geolocated Twitter social media data to describe the geographic spread of SARS-CoV-2. **Journal of travel medicine**, 27(5), 2020.

CACERES, N.; WIDEBERG, J. P.; BENITEZ, F. GARCIA. Review of traffic data estimations extracted from cellular networks. **IET Intelligent Transport Systems**, v. 2, n. 3, p. 179-192, 2008.

CARVALHOA, A.; FERREIRA, M. C.; DIAS, T. G. **Understanding mobility patterns and user activities from geo-tagged social networks data**, 2020.

CASCETTA, E. Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. **Transportation Research Part B: Methodological**, v. 18, n. 4, p. 289-299, 1984.

CEBELAK, M. K. **Transportation planning via location-based social networking data: exploring many-to-many connections**. (Doctoral dissertation), 2015.

CERRONE, D., PAU, H.; LEHTOVUORI, P. **A sense of place. Exploring the potentials and possible uses of Location Based Social Network Data for urban and transportation planning in Turku City Centre**, 2015.

CERRONE, D.; BAEZA, J. L.; LEHTOVUORI, P. Optional and necessary activities: operationalising Jan Gehl's analysis of urban space with Foursquare data. **International Journal of Knowledge-Based Development**, 11(1), p. 68-79, 2020.

CHANIOTAKIS, E.; ANTONIOU, C. **Use of geotagged social media in urban settings: Empirical evidence on its potential from Twitter**. International Conference on Intelligent Transportation Systems, IEEE, p. 214-219, 2015.

CHANIOTAKIS, E.; ANTONIOU, C.; GRAU, J. M. S.; DIMITRIOU, L. **Can Social Media data augment travel demand survey data?** International Conference on Intelligent Transportation Systems, IEEE, p. 1642-1647, 2016.

CHEN, Y., HU, J., XIAO, Y., LI, X., HUI, P. Understanding the User Behavior of Foursquare: A Data-Driven Study on a Global Scale. **IEEE Transactions on Computational Social Systems**, 2020.

CHENG, Z.; CAVERLEE, J.; LEE, K.; SUI, D. Z. **Exploring millions of footprints in location sharing services**. In Fifth International AAAI Conference on Weblogs and Social Media, 2011.

CHENG, Z.; JIAN, S. MAGHREBI, M.; RASHIDI, T.; WALLER, S. Is Social Media an Appropriate Data Source to Improve Travel Demand Estimation Models? **Transportation Research Board**, 2018.

CRANSHAW, J.; SCHWARTZ, R.; HONG, J.; E SADEH, N. **The livehoods project: Utilizing social media to understand the dynamics of a city.** In **Sixth International AAAI Conference on Weblogs and Social Media**, 2012.

DANIELE, Q.; AIELLO, L. M.; SCHIFANELLA, R.; ADAM, D. **The digital life of walkable streets.** In 24th International Conference on World Wide Web, p. 875-884, 2015.

DEEVA, G.; SMEDT, J.; WEERDT, J.; OSKARSDOTTIR, M. **Mining Behavioural Patterns in Urban Mobility Sequences Using Foursquare Check-in Data from Tokyo.** In International Conference on Complex Networks and Their Applications (pp. 931-943). Springer, 2019.

DEVILLE, P.; LINARD, C.; MARTIN, S.; GILBERT, M.; STEVENS, F. R.; GAUGHAN, A. E.; TATEM, A. J. Dynamic population mapping using mobile phone data. **Proceedings of the National Academy of Sciences**, 111(45), 15888-15893, 2014.

DHARMOWIJOYO, D. B.; SUSILO, Y. O.; KARLSTROM, A.; ADIREDDA, L. S. Collecting a multi-dimensional three-weeks household time-use and activity diary in the Bandung Metropolitan Area, Indonesia. **Transportation Research Part A: Policy and Practice**, v. 80, p. 231-246, 2015.

EBRAHIMPOUR, Z.; WAN, W.; VELÁZQUEZ GARCÍA, J. L.; CERVANTES, O.; HOU, L. Analyzing Social-Geographic Human Mobility Patterns Using Large-Scale Social Media Data. **ISPRS International Journal of Geo-Information**, 9(2), 125, 2020.

EGU, O.; BONNEL, P. How comparable are origin-destination matrices estimated from automatic fare collection, origin-destination surveys and household travel survey? An empirical investigation in Lyon. **Transportation Research Part A: Policy and Practice**, 138, 267-282, 2020.

ELANGO, VETRI V.; GUENSLER, RANDALL; OGLE, J. Day-to-day travel variability in the Commute Atlanta, Georgia, study. **Transportation research record**, n. 1, p. 39-49, 2007.

ELLIS, K.; GOLDSZMIDT, M.; LANCKRIET, G.; MISHRA, N.; REINGOLD, O. **Equality and social mobility in Twitter discussion groups.** In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, p. 523-532, 2016.

FAGNANT, D. J.; KOCKELMAN, K. M. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. **Transportation Research Part C: Emerging Technologies**, v. 40, p. 1-13, 2014.



FAYE, S.; BRONZI, W.; TAHIROU, I.; ENGEL, T. Characterizing user mobility using mobile sensing systems. **International Journal of Distributed Sensor Networks**, 13(8), 2017.

FERREIRA, A. P. G.; SILVA, T. H.; LOUREIRO, A. A. F. **Beyond sights: Large scale study of tourists' behavior using Foursquare data.** In 2015 IEEE International Conference on Data Mining Workshop (ICDMW), p. 1117-1124, 2015.

FERREIRA, J. **Mobi-System: towards an information system to support sustainable mobility with electric vehicle integration.** Doctoral thesis. Minho's University (Portugal), 2013.

FORGHANI, M.; KARIMIPOUR, F.; CLARAMUNT, C. From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. **Transportation Research Part C: Emerging Technologies**, v. 117, p. 102666, 2020.

FRITH, J. Turning life into a game: Foursquare, gamification, and personal mobility. **Mobile Media & Communication**, v. 1, n. 2, p. 248-262, 2013.

GAL-TZUR, A.; GRANT-MULLER, S. M.; KUFLIK, T.; MINKOV, E.; NOCERA, S., SHOOR, I. **The potential of social media in delivering transport policy goals.** *Transport Policy*, 32, 115-123, 2014.

GAO, S.; YANG, J. A.; YAN, B.; HU, Y.; JANOWICZ, K.; MCKENZIE, G. **Detecting origin-destination mobility flows from geotagged tweets in greater Los Angeles area.** In Eighth Intl. Conference on Geographic Information Science, 2014.

GRABOWICZ, P. A.; RAMASCO, J. J.; GONÇALVES, B.; EGUÍLUZ, V. M. **Entangling mobility and interactions in social media.** *PloS one*, 9(3), e92196, 2014.

GRAHAM, M.; HALE, S. A.; GAFFNEY, D. Where in the world are you? Geolocation and language identification in Twitter. **The Professional Geographer**, 66(4), 568-578, 2014.

GROVES, R. M. **Nonresponse rates and nonresponse bias in household surveys.** *Public opinion quarterly*, 70(5), p. 646-675, 2006.

GU, Y.; QIAN, Z. S.; CHEN, F. From Twitter to detector: Real-time traffic incident detection using social media data. **Transportation research part C: emerging technologies**, 67, 321-342, 2016.

HANSON, S.; HUFF, J. O. Assessing day-to-day variability in complex travel patterns. **Transportation Research Record**, v. 891, p. 18-24, 1981.

HASAN, S.; ZHAN, X.; E UKKUSURI, S. V. **Understanding urban human activity and mobility patterns using large-scale location-based data from online social media.** In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing, pp. 1-8, 2013.

HAWELKA, B.; SITKO, I.; BEINAT, E.; SOBOLEVSKY, S.; KAZAKOPOULOS, P.; RATTI, C. Geo-located Twitter as proxy for global mobility patterns. **Cartography and Geographic Information Science**, v. 41, n. 3, p. 260-271, 2014.

HAZELTON, M. L. Some comments on origin–destination matrix estimation. **Transportation Research Part A: Policy and Practice**, 37(10), 811-822, 2008.

HELLINGA, B. **Estimating dynamic origin-destination demands from link and probe counts**. Queen's University at Kingston, 1995.

HONG, L.; AHMED, A.; GURUMURTHY, S.; SMOLA, A. J.; TSIOUTSIOLIKLIS, K. **Discovering geographical topics in the Twitter stream**. In Proceedings of the 21st international conference on World Wide Web p.769-778, 2012.

HSIAO, Y.; FIORIO, L.; WAKEFIELD, J.; ZAGHENI, E. **Modeling the bias of digital data: an approach to combining digital and survey data to estimate and predict migration trends**. Max Planck Institute for Demographic Research, Rostock, Germany, 2020.

HU, L.; LI, Z.; YE, X. Delineating and modeling activity space using geotagged social media data. **Cartography and Geographic Information Science**, 47(3), p. 277-288, 2020.

HU, W.; JIN, P. J. An adaptive hawkes process formulation for estimating time-of-day zonal trip arrivals with location-based social networking check-in data. **Transportation Research Part C: Emerging Technologies**, v. 79, p. 136-155, 2017.

HU, W.; JIN, P. J. **Dynamic Origin–Destination Estimation Based on Time-Delay Correlation Analysis on Location-Based Social Network Data**, 2018.

HUANG, Q.; WONG, D. W. **Modeling and visualizing regular human mobility patterns with uncertainty: An example using Twitter data**. Annals of the Association of American Geographers, 105(6), p. 1179-1197, 2015.

IQBAL, M. S.; CHOUDHURY, C. F.; WANG, P.; GONZALEZ, M. C. Development of origin – destination matrices using mobile phone call data. **Transportation Research Part C: Emerging Technologies**, v. 40, p. 63-74, 2014.

JARV, OLLE; AHAS, REIN; WITLOX, FRANK. 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, p. 122-135, 2014.

JIANG, B.; THILL, J. C. Volunteered Geographic Information: Towards the establishment of a new paradigm. **Computers, Environment and Urban Systems – Journal**, v. 53. 2015.

JIANG, SHAN; FERREIRA, JOSEPH; GONZALEZ, MARTA C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. **IEEE Transactions on Big data**, v. 3, n. 2, p. 208-219, 2017.

- JIANG, Y.; LI, Z.; YE, X. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. **Cartography and Geographic Information Science**, v. 46, n. 3, p. 228-242, 2019.
- JIN, P. J.; CEBELAK, M.; YANG, F.; ZHANG, J.; WALTON, C. M.; RAN, B. Location-based social networking data: Exploration into use of doubly constrained gravity model for origin–destination estimation. **Transportation Research Record**, 2430(1), 72-82, 2014.
- JONES, A. S.; GEORGAKIS, P.; PETALAS, Y.; SURESH, R. **Real-time traffic event detection using Twitter data**. *Infrastructure Asset Management*, 5(3), p. 77-84, 2018.
- JONES, P.; CLARKE, M. **The significance and measurement of variability in travel behaviour**. *Transportation*, v. 15, n. 1-2, p. 65-87, 1988.
- JURDAK, R.; ZHAO, K.; LIU, J.; ABOUJAOUDE, M.; CAMERON, M.; NEWTH, D. **Understanding human mobility from Twitter**. *PloS one*, 10(7), e0131469, 2015.
- KHAN, S. F.; BERGMANN, N.; JURDAK, R.; KUSY, B.; CAMERON, M. **Mobility in cities: Comparative analysis of mobility models using Geo-tagged tweets in Australia**. In 2017 IEEE 2nd International Conference on Big data Analysis (ICBDA), p. 816-822, 2017.
- KHEIRI, A.; KARIMIPOUR, F.; FORGHANI, M. **Intra-urban movement flow estimation using location based social networking data**. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(1), 781, 2015.
- KOKKINOGENIS, Z.; FILGUIERAS, J.; CARVALHO, S.; SARMENTO, L.; ROSSETTI, R. J. **Mobility network evaluation in the user perspective: Real-time sensing of traffic information in Twitter messages**. In *Advances in Artificial Transportation Systems and Simulation*, p. 219-234, 2015.
- KOPPELMAN, F. S.; PAS, E. I. **Estimation of disaggregate regression models of person trip generation with multiday data**. Papers presented during the Ninth International Symposium on Transportation and Traffic Theory held in Delft the Netherlands, p. 11-13, 1984.
- KRAEMER, M. U.; BISANZIO, D.; REINER, R. C.; ZAKAR, R.; HAWKINS, J. B.; FREIFELD, C. C.; PERKINS, T. A. **Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about human mobility: a case study using geolocated tweets from Lahore, Pakistan**. *EPJ Data Science*, 2018.
- KUNG, K. S.; GRECO, K.; SOBOLEVSKY, S.; RATTI, C. **Exploring universal patterns in human home-work commuting from mobile phone data**. *PloS one*, 9(6), e96180, 2014.
- KURKCU, A.; OZBAY, K.; MORGUL, E. F. **Evaluating the usability of geo-located Twitter as a tool for human activity and mobility patterns: A case study for nyc**. In *Transportation Research Board's 95th Annual Meeting*, p. 1-20, 2016.

LENORMAND, M.; PICORNELL, M.; CANTU-ROS, O. G.; TUGORES, A.; LOUAIL, T.; HERRANZ, R.; RAMASCO, J. J. **Cross-checking different sources of mobility information.** *PloS one*, 9(8), 2014.

LI, L.; GOODCHILD, M. F.; XU, B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. **Cartography and geographic information science**, v. 40, n. 2, p. 61-77, 2013.

LI, M.; WESTERHOLT, R.; FAN, H.; ZIPF, A. **Assessing spatiotemporal predictability of LBSN: a case study of three Foursquare datasets.** *Geo Informatica*, 22(3), 541-561, 2018.

LI, Y.; LI, Q.; SHAN, J. Discover patterns and mobility of Twitter users—A study of four US college cities. **ISPRS International Journal of Geo-Information**, 6(2), 42, 2017.

LIAO, Y.; YEH, S.; JEUKEN, G. S. **From individual to collective behaviours: exploring population heterogeneity of human mobility based on social media data.** *EPJ Data Science*, 8(1), 34, 2019.

LIMA, L. S. **Espraçamento Urbano por Autossegregação e seus Impactos na Acessibilidade Urbana de Fortaleza.** Dissertação de mestrado. Universidade Federal do Ceará, Fortaleza – CE, 2017.

LIU, J.; ZHAO, K.; KHAN, S.; CAMERON, M.; JURDAK, R. **Multi-scale population and mobility estimation with geo-tagged tweets.** In 2015 31st IEEE International Conference on Data Engineering Workshops, p. 83-86, 2015.

LLORENTE, A.; GARCIA-HERRANZ, M.; CEBRIAN, M.; MORO, E. **Social media fingerprints of unemployment.** *PloS one*, 10(5), 2015.

LUCERI, L.; ANDREOLETTI, D.; TORNATORE, M.; BRAUN, T.; GIORDANO, S. **Measurement and control of geo-location privacy on Twitter.** *Online social networks and media*, 17, 2020.

LUO, F.; CAO, G.; MULLIGAN, K.; LI, X. **Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago.** *Applied Geography*, 70, 11-25, 2016.

MARBLE, D. F.; BOWLBY, S. R. **Shopping alternatives and recurrent travel patterns.** *Geographic studies of urban transportation and network analysis*, p. 42-75, 1968.

MARTI, P., SERRANO-ESTRADA, L., & NOLASCO-CIRUGEDA, A. **Social media data: Challenges, opportunities and limitations in urban studies.** *Computers, Environment and Urban Systems*, 74, 161-174, 2019.

MARTI, P.; SERRANO-ESTRADA, L.; NOLASCO-CIRUGEDA, A. **Using locative social media and urban cartographies to identify and locate successful urban plazas.** *Cities*, 64, p. 66-78, 2017.

MARTIN, Y.; CUTTER, S. L.; LI, Z. **Bridging Twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma.** *Natural hazards review*, 21(2), 2020.

MAZZOLI, M.; DIECHTIAREFF, B.; TUGORES, A.; WIVES, W.; ADLER, N.; COLET, P.; RAMASCO, J. J. **Migrant mobility flows characterized with digital data.** *Plos one*, 15(3), 2020.

MICHAU, G.; NANTES, A.; CHUNG, E.; ABRY, P.; BORGNAT, P. **Retrieving dynamic Origin-Destination matrices from bluetooth data**, 2014.

MIRZAEI, S.; WANG, Q. **Urban mobility and resilience: exploring Boston's urban mobility network through Twitter data.** *Applied Network Science*, 5(1), 1-20, 2020.

MIYAZAWA, S.; SONG, X.; JIANG, R.; FAN, Z.; SHIBASAKI, R.; SATO, T. **City-Scale Human Mobility Prediction Model by Integrating Gns Trajectories and Sns Data Using Long Short-Term Memory.** *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 5(4), 2020.

MIYAZAWA, S.; SONG, X.; XIA, T.; SHIBASAKI, R.; KANEDA, H. **Integrating GPS trajectory and topics from Twitter stream for human mobility estimation.** *Frontiers of Computer Science*, 13(3), 460-470, 2019.

MOHAMED, K.; COME, E.; OUKHELLOU, L.; VERLEYSSEN, M. **Clustering smart card data for urban mobility analysis.** *IEEE Transactions on intelligent transportation systems*, 18(3), 712-728, 2016.

MOLLOY, J.; MOECKEL, R. **Improving destination choice modeling using location-based Big data.** *ISPRS International Journal of Geo-Information*, v. 6, n. 9, p. 291, 2017.

MOTAHARI, S.; ZANG, H.; REUTHER, P. **The impact of temporal factors on mobility patterns.** In: 2012 45th Hawaii International Conference on System Sciences. IEEE, p. 5659-5668, 2012.

NI, M.; HE, Q.; GAO, J. **Forecasting the subway passenger flow under event occurrences with social media.** *IEEE Transactions on Intelligent Transportation Systems*, 18(6), 1623-1632, 2016.

NIKOLAIDOU, A.; PAPAIOANNOU, P. **Utilizing social media in transport planning and public transit quality: Survey of literature.** *Journal of Transportation Engineering, Part A: Systems*, 144(4), 2018.

NOVOVIC, O.; GRUJIC, N.; BRDAR, S.; GOVEDARICA, M.; CRNOJEVIC, V. **Clustering Foursquare Mobility Networks to Explore Urban Spaces.** In *World Conference on Information Systems and Technologies*, p. 544-553, 2020.

OLIVEIRA, M. V. **A natureza dos padrões de variação espaço-temporal do volume veicular em ambiente urbano: estudo de caso em Fortaleza.** Programa de Pós-

graduação em Engenharia de Transportes–PETRAN. Universidade Federal do Ceará, Fortaleza, 2004.

OLIVEIRA, S.F.C.; SILVA C. A. U.; SILVA, M. Y. O. O uso do Instagram e do Twitter para identificação e mapeamento de usos do solo em áreas urbanas: o caso de Fortaleza-CE. **Brazilian Journal of Development**, v. 5, n. 9, p. 16128-16149, 2019.

OMRI, M. **Every Tweet Counts: Examining Spatial Variability of Twitter Data Representativeness**. Master Thesis, Graduate School of the University of Oregon, 2016.

OPOVO. **Mais de 23 mil domicílios serão visitados para pesquisa Origem-Destino**. Disponível em: <https://www.opovo.com.br/noticias/fortaleza/2019/01/28638-mais-de-23-mil-domicilios-serao-visitados-para-pesquisa-origem-destino.html>. Acessado em: 15/02/2020.

PAJEVIC, F.; SHEARMUR, R. G. Catch me if you can: Workplace mobility and Big data. **Journal of Urban Technology**, v. 24, n. 3, p. 99-115, 2017.

PAS, E. I. Multiday samples, parameter estimation precision, and data collection costs for least squares regression trip-generation models. **Environment and Planning A**, v. 18, n. 1, p. 73-87, 1986.

PAS, E. I.; SUNDAR, S. Intrapersonal variability in daily urban travel behavior: some additional evidence. **Transportation**, v. 22, n. 2, p. 135-150, 1995.

PASZTO, V.; BURIAN, J.; MACKU, K. Changing Mobility Lifestyle: A Case Study on the Impact of COVID-19 Using Personal Google Locations Data. **International Journal of E-Planning Research**, 10(2), 66-79, 2020.

PAULE, J. D. G.; SUN, Y.; MOSHFEGHI, Y. **On fine-grained geolocalisation of tweets and real-time traffic incident detection**. *Information Processing & Management*, 56(3), 1119-113, 2019.

PENDYALA, R. M.; PAS, E. I. Multi-day and multi-period data for travel demand analysis and modeling. **Transportation Research Circular**, 2000.

PEREIRA, R. H. **Tempo de deslocamento casa-trabalho no Brasil (1992-2009): diferenças entre regiões metropolitanas, níveis de renda e sexo**. Texto para Discussão, Instituto de Pesquisa Econômica Aplicada (IPEA), 2013.

PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS CONTÍNUA (IBGE). **Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento**, 2018.

PIMPALÉ, P.; PANANGADAN, A.; ABELLERA, L. V. **Analyzing spread of influence in social networks for transportation applications**. In 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), p. 763-768, 2018.

PIRES, A. B.; VASCONCELLOS, E. A.; SILVA, A. Camargo e. **Transporte humano: cidades com qualidade de vida**. São Paulo: ANTP, Brasil, 1997.

PIZA, M. **O fenômeno Instagram: considerações sob a perspectiva tecnológica.** Monografia (Graduação em Sociologia). Universidade de Brasília, Brasília, 2012.

POUREBRAHIM, N.; SULTANA, S.; NIAKANLAHIJI, A.; THILL, J. C. Trip distribution modeling with Twitter data. **Computers, Environment and Urban Systems**, v. 77, p. 101354, 2019.

POUREBRAHIM, N.; SULTANA, S.; THILL, J. C.; MOHANTY, S. **Enhancing trip distribution prediction with Twitter data: comparison of neural network and gravity models.** In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery. 2018. p. 5-8, 2018.

PREFEITURA MUNICIPAL DE FORTALEZA, **Fortaleza 2040. Cidade Conectada, Acessível e Justa. Plano de Mobilidade e Acessibilidade Urbano.** Vol. 3. Parte I, II e III, 2015.

PROVENZANO, DAVIDE; HAWELKA, BARTOSZ; BAGGIO, RODOLFO. The mobility network of European tourists: a longitudinal study and a comparison with geo-located Twitter data. **Tourism Review**, 2018.

PU, Z.; ZHU, M.; LI, W.; CUI, Z.; GUO, X.; WANG, Y. **Monitoring Public Transit Ridership Flow by Passively Sensing Wi-Fi and Bluetooth Mobile Devices.** IEEE Internet of Things Journal, 2020.

QUERCIA, D.; AIELLO, L. M.; SCHIFANELLA, R.; DAVIES, A. **The digital life of walkable streets.** In Proceedings of the 24th international conference on World Wide Web, p. 875-884, 2015.

QUERCIA, D.; DI LORENZO, G.; CALABRESE, F.; RATTI, C. **Mobile phones and outdoor advertising: measurable advertising.** Institute of Electrical and Electronics Engineers, 2011.

RAMOS, F. R. **Análise espacial de estruturas intra-urbanas: o caso de São Paulo.** São José dos Campos: INPE, 2002.

RASHIDI, T. H.; ABBASI, A.; MAGHREBI, M.; HASAN, S.; WALLER, T. S. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. **Transportation Research Part C: Emerging Technologies**, 75, pp. 197-211, 2017.

RAUX, C.; MA, T.; CORNELIS, E. Variability in daily activity-travel patterns: the case of a one-week travel diary. **European transport research review**, v. 8, n. 4, p. 26, 2016.

RIBEIRO, A. I. J. T.; SILVA, T. H.; DUARTE-FIGUEIREDO, F.; LOUREIRO, A. A. **Studying traffic conditions by analyzing Foursquare and Instagram data.** In Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks, p. 17-24, 2014.

ROBERTSON, C.; FEICK, R. Bumps and bruises in the digital skins of cities: Unevenly distributed user-generated content across US urban areas. **Cartography and Geographic Information Science**, v. 43, n. 4, p. 283-300, 2016.

ROBILLARD, P. Estimating the OD matrix from observed link volumes. **Transportation Research**, 9(2-3), p. 123-128, 1975.

SALAS, O., M. H., QUEZADA, C. R. The use of public spaces in a medium-sized city: from Twitter data to mobility patterns. **Journal of Maps**, 13(1), pp. 40-45, 2017.

SCHLICH, R.; AXHAUSEN, K. W. Habitual travel behaviour: evidence from a six-week travel diary. **Transportation**, v. 30, n. 1, p. 13-36, 2003.

SHAPCOTT, M. **Comparison of the use of time in Reading, England with time use in other countries**. Transactions of the Martin Centre for Architectural and Urban Studies, v. 3, p. 231-257, 1978.

SLOAN, L.; QUAN-HAASE, A. **The SAGE handbook of social media research methods**. Sage, 2017.

SOLIMAN, A.; SOLTANI, K.; YIN, J.; PADMANABHAN, A.; WANG, S. **Social sensing of urban land use based on analysis of Twitter users' mobility patterns**. PloS one, 12(7), 2017.

SOUSA, F. F. L. D. M.; MESQUITA, K. G. A.; LOUREIRO, C. F. G. Caracterização da evolução do padrão de mobilidade de fortaleza a partir da calibração do tranus. In: **XIX ANPET Congresso Nacional de Pesquisa e Ensino em Transportes, 2019**, Balneário Camboriú. Anais do XXXI ANPET Congresso Nacional de Pesquisa e Ensino em Transportes, 2019.

STOPHER, P. R.; ZHANG, Y. Repetitiveness of daily travel. **Transportation Research Record**, v. 2230, n. 1, p. 75-84, 2011.

SULIS, P., MANLEY, E., ZHONG, C., & BATTY, M. Using mobility data as proxy for measuring urban vitality. **Journal of Spatial Information Science**, 137-162, 2018.

SUN, Y.; LI, M. Investigation of travel and activity patterns using location-based social network data: A case study of active mobile social media users. **ISPRS International Journal of Geo-Information**, v. 4, n. 3, p. 1512-1529, 2015.

SUSILO, Y. O.; KITAMURA, R. Analysis of day-to-day variability in an individual's action space: exploration of 6-week Mobidrive travel diary data. **Transportation Research Record**, v. 1902, n. 1, p. 124-133, 2005.

TARIGAN, A. K.; KITAMURA, R. Week-to-week leisure trip frequency and its variability. **Transportation Research Record**, v. 2135, n. 1, p. 43-51, 2009.

TASSE, D.; HONG, J. I. Using user-generated content to understand cities. In *Seeing cities through Big data*. Springer, Cham, p. 49-64, 2017.



VERSCHELE, M.; NEUTENS, T.; DELAFONTAINE, M.; VAN DE WEGHE, N. **The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities.** *Applied Geography*, 32(2), p. 208-220, 2012.

VIEIRA, M. R.; FIGUEIREDO, J. M. D.; LIBERATTI, G.; VIEBRANTZ, A. F. M. **Bancos de Dados NoSQL: conceitos, ferramentas, linguagens e estudos de casos no contexto de Big data.** *Simpósio Brasileiro de Bancos de Dados*, 27, 1-30, 2012.

WANG, F.; CHEN, C. On data processing required to derive mobility patterns from passively-generated mobile phone data. **Transportation Research Part C: Emerging Technologies**, v. 87, p. 58-74, 2018.

WANG, Yan; TAYLOR, John E. Coupling sentiment and human mobility in natural disasters: a Twitter-based study of the 2014 South Napa Earthquake. **Natural hazards**, v. 92, n. 2, p. 907-925, 2018.

WANICHAYAPONG, N.; PRUTHIPUNYASKUL, W.; PATTARA-ATIKOM, W.; CHAOVALIT, P. **Social-based traffic information extraction and classification.** In 2011 11th International Conference on ITS Telecommunications, p. 107-112, 2011.

WOLF, J. L. **Using GPS data loggers to replace travel diaries in the collection of travel data.** (Doctoral dissertation, School of Civil and Environmental Engineering, Georgia Institute of Technology), 2000.

WOLF, J.; OLIVEIRA, M.; THOMPSON, M. Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey. **Transportation Research Record: Journal of the Transportation Research Board**, n.1854, p. 189-198, 2003.

WRIGHT, T.; HU, P.; YOUNG, J.; LU, A. **Variability in Traffic Monitoring Data: Final Summary Report.** US Department of Energy, Washington, DC, 1997.

WU, H.; LIU, L.; YU, Y.; PENG, Z.; JIAO, H.; NIU, Q. An agent-based model simulation of human mobility based on mobile phone data: how commuting relates to congestion. **ISPRS International Journal of Geo-Information**, 8(7), 2019.

XIANYU, J.; RASOULI, S.; TIMMERMANS, H. Analysis of variability in multi-day GPS imputed activity-travel diaries using multi-dimensional sequence alignment and panel effects regression models. **Transportation**, v. 44, n. 3, p. 533-553, 2017.

XU, S.; LI, S.; WEN, R.; HUANG, W. **Traffic event detection using Twitter data based on association rules.** *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4, 2019.

YANG, F.; JIN, P. J.; CHENG, Y.; ZHANG, J.; RAN, B. Origin-destination estimation for non-commuting trips using location-based social networking data. **International Journal of Sustainable Transportation**, 9(8), 551-564, 2015.

YANG, Liya; DURARTE, Carlos Marmolejo. Identifying tourist-functional relations of urban places through Foursquare from Barcelona. **GeoJournal**, p. 1-18, 2019.

YIGITCANLAR, T.; KANKANAMGE, N.; VELLA, K. How Are Smart City Concepts and Technologies Perceived and Utilized? A Systematic Geo-Twitter Analysis of Smart Cities in Australia. **Journal of Urban Technology**, 1-20, 2020.

YIN, J.; GAO, Y.; DU, Z.; WANG, S. Exploring multi-scale spatiotemporal Twitter user mobility patterns with a visual-analytics approach. **ISPRS International Journal of Geo-Information**, 5(10), 187, 2016.

YIN, J.; SOLIMAN, A.; YIN, D.; WANG, S. Depicting urban boundaries from a mobility network of spatial interactions: a case study of Great Britain with geo-located Twitter data. **International Journal of Geographical Information Science**, v. 31, n. 7, p. 1293-1313, 2017.

YOO, B. S.; KANG, S. P.; CHON, K.; KIM, S. G. Origin-destination estimation using cellular phone BS information. **Journal of the Eastern Asia Society for Transportation Studies**, 6, 2574-2588, 2005.

YUAN, Y.; MILLS, D. (2019). **Exploring Urban Dynamics from Bluetooth Tracking Data: A Case Study of Austin, Texas**. Proc. Int. Cartogr. Assoc, 2, 153, 2019.

ZHANG, Z.; NI, M., HE, Q.; GAO, J.; GOU, J.; LI, X. Exploratory study on correlation between Twitter concentration and traffic surges. **Transportation Research Record**, 2553(1), 90-98, 2016.

ZHONG, C.; MANLEY, E.; ARISONA, S. M.; BATTY, M.; SCHMITT, G. Measuring variability of mobility patterns from multiday smart-card data. **Journal of Computational Science**, v. 9, p. 125-130, 2015.

## ANEXO

1: Rotina de coleta dos dados do *Twitter*

```

1 from twython import Twython#, TwythonStreamer
2 import json
3 import time
4 import datetime
5
6 now = datetime.datetime.now()
7 arq1 = './dados/arquivo.txt' % (now.year, now.month, now.
  day, now.hour, now.minute)
8 arq2 = './dados/arquivo1.csv' % (now.year, now.month, now.
  day, now.hour, now.minute)
9 hh = now.hour
10 resultado=open(arq1,"w")
11 resultadocsv=open(arq2,"w")
12
13 # leitura dos centroides (áreas escolhidas);
14
15 arqpontos=open("pontoscentroides.csv","r")
16
17 cabecalho=arqpontos.readline()
18
19 # Dimensionamento de 2 Km
20
21 geocode=[]
22 for x in arqpontos.readlines():
23     p=x.split(";")
24     r = 2
25     geocode.append(p[1].strip() +"," + p[0] +"," + str(r)+ "
  km")
26
27 with open('./conf/app.json', 'r') as app_file:
28     app = json.load(app_file)
29
30 # Twitter Rest API (Chaves de acesso disponibilizadas em: '
  https://developer.twitter.com/en/user/login')
31 twitter = Twython(app["CONSUMER_KEY"], app["
  CONSUMER_SECRET"], oauth_version=2)
32 ACCESS_TOKEN = twitter.obtain_access_token()
33
34 twitter = Twython(app["CONSUMER_KEY"], access_token=
  ACCESS_TOKEN)
35
36 t = 30 * 24 * 3600
37 inicio=time.clock_gettime(0)
38 contador=0
39
40 # Informações obtidas
41

```

```

42 resultadocsv.write("idUsuario;DataeHora;Nome;texto;Fonte;
Latitude;Longitude\n")
43
44 # Processos adicionais para a coleta
45
46 while((time.clock_gettime(0)-inicio)<=t):
47     # Other queries
48     now = datetime.datetime.now()
49     nhh = now.hour
50     dia = int((time.clock_gettime(0)-inicio)/86400)+1
51     if nhh != hh:
52         resultado.close()
53         resultadocsv.close()
54         arq1 = './dados/arq%4d%02d%02d_%02d_%02d.txt' % (
now.year, now.month, now.day, now.hour, now.minute)
55         arq2 = './dados/arq%4d%02d%02d_%02d_%02d.csv' % (
now.year, now.month, now.day, now.hour, now.minute)
56         resultado=open(arq1,"w")
57         resultadocsv=open(arq2,"w")
58         print(arq1)
59         hh = nhh
60
61     for g in geocode:
62         if (time.clock_gettime(0)-inicio)>=t:
63             break
64         try:
65             geo_located_tweets = twitter.search(geocode=g
, count=10, result_type="recent")
66
67             tweet = geo_located_tweets[u'statuses'][0]
68             #print(json.dumps(tweet, indent=2))
69             hora=tweet["created_at"]
70             id2usuario=tweet["user"]["id"]
71             nomedousuario=tweet["user"]["name"]
72             texto=tweet["text"]
73             fonte=tweet["source"]
74             if tweet["geo"] !=None:
75                 latitude=tweet["geo"]["coordinates"][0]
76                 longitude=tweet["geo"]["coordinates"][1]
77
78 # Principais informações a serem
79
80             resultadocsv.write(str(id2usuario)+";"+
hora+";"+(nomedousuario)+";"+(texto)+";"+(fonte)+";"+str(
latitude)+";"+str(longitude)+"\n")
81
82             print(dia,hh,tweet["geo"])

```

```
83         resultado.write(json.dumps(tweet, indent=2))
84         resultado.write("\n")
85         contador=contador+1
86         time.sleep(2)
87     except:
88         print("*")
89         continue
90
91 print(contador)
92 resultado.close()
93 resultadocsv.close()
94
95
96
```

## 2: Rotina para estruturação, classificação e obtenção da matriz OD

```
1 # Construção de fluxos entre id's
2
3 require(data.table)
4 require(dplyr)
5 require(lubridate)
6 require(tidyr)
7 require(ggplot2)
8
9 setwd('~/.diretorio/')
10 pasta <- "~/.diretorio"
11
12 tmp0 <- fread(paste(pasta, "arquivo.csv", sep=''), header
  =TRUE)
13
14 # Estruturação da data de cada check-in
15
16 tmp0$data <- dmy_hm(tmp0$data)
17 tmp0$dia <- day(tmp0$data)
18 tmp0$mes <- month(tmp0$data)
19 tmp0$ano <- year(tmp0$data)
20 tmp0$ds <- wday(tmp0$data)
21 tmp0$hora <- hour(tmp0$data)
22
23 # Classificando as postagens por horário
24 # T = Trabalho
25 # C = Casa
26
27 l <- function(x)
28 {
29   if (is.element(x, c(7:18))) {
30     l <- 'T'
31   } else {
32     l <- 'C'
33   }
34   return (l)
35 }
36
37
38 # Todas as postagens
39 tmp0 <- arrange(tmp0, data)
40
41 # Descobrimos os usuários com duas ou mais postagens
42 tmp1 <- data.table(tmp0) %>% group_by(id) %>%
  summarise(n=n()) %>% filter(n >= 2)
43
44 # Filtrando as postagens para captar apenas os usuários
  com duas ou mais postagens
```

```

45 tmp2 <- data.table(tmp0) %>% filter(id % in % tmp1$id,
    bairro != 226)
46
47 # Colocar o local nas postagens
48 tmp2$local <- sapply(tmp2$hora, l)
49
50 # Número de postagens em casa e no trabalho
51 tmp3 <- group_by(tmp2, id) %>%
52 summarise(total=n(),
53 PC = sum(local == 'C'),
54 PT = sum(local == 'T'))
55
56 # Filtrar PT ou PC = 0
57 # PT = postagens casa
58 # PC = Postagens trabalho
59 tmp4 <- filter(tmp3, !((PT * PC) == 0))
60
61 # Filtrando as postagens para pegar apenas os usuários com
    postagens T e C
62 tmp5 <- filter(tmp2, id % in % tmp4$id)
63
64 # Criando as listas de PCs e PTs para cada usuário
65 tmp6 <- group_by(tmp5, id) %>%
66 summarise(total=n(),
67 PC = sum(local == 'C'),
68 PT = sum(local == 'T'),
69 NB = length(unique(bairro)),
70 ColPC = list(sort(bairro[local == "C"])),
71 ColPT = list(sort(bairro[local == "T"])),
72 UniPC = list(sort(unique(bairro[local == "C"]))),
73 UniPT = list(sort(unique(bairro[local == "T"])))) %>%
74
75 filter(NB > 1)
76
77 tmp6$ColPC <- tmp6$ColPC
78 for (row in 1:nrow(tmp6)){
79 colx <- unlist(tmp6[row, 'ColPC'], use.names = F)
80 coly <- unlist(tmp6[row, 'UniPT'], use.names = F)
81 if ((length(coly) == 1) && (length(colx) > 1)) {
82 if (any(colx % in % coly)) {
83
84 tmp6$ColPC[row] <- list(colx[-which(colx % in % coly)])
85
86 }
87 }
88 }
89

```

```

90 tmp6$nColPT <- tmp6$ColPT
91 for (row in 1:nrow(tmp6)){
92 tmp6$PrPC[row] <- names(sort(table(tmp6$nColPC[row]),
  decreasing = T)[1])
93 colx <- unlist(tmp6[row, 'ColPT'], use.names = F)
94 coly <- unlist(tmp6[row, 'PrPC'], use.names = F)
95 if (length(colx) > 1) {
96 if (any(colx % in % coly)) {
97 tmp6$nColPT[row] <- list(colx[-which(colx % in % coly)])
98 }
99 }
100 }
101
102 # Filtrando as postagens para captar apenas os usuários
  com viagens completas
103 tmp7c <- filter(tmp5, id % in % tmp6$id, local == 'C')
  # para origens
104 tmp7t <- filter(tmp5, id % in % tmp6$id, local == 'T')
  # para destinos
105 origens <- select(tmp6, id, PrPC)
106 destinos <- select(tmp6, id, nColPT)
107 colnames(origens) <- c('id', 'bairro')
108
109 # Criando a lista de origens de viagens
110 origens <- mutate(origens, bairro=as.numeric(bairro))
111 tmp8 <- left_join(origens, tmp7c, by=c('id', 'bairro'
  )) % > %
112 filter(!duplicated(id))
113
114 # Criando uma tabela vazia para os destinos
115 ndest <- data.table(
116 id = character(),
117 bairro = integer()
118 )
119
120 # Criando a lista de destinos de viagens
121 for (row in 1:nrow(destinos)){
122 xid <- unlist(destinos[row, 'id'], use.names = F)
123 xdest <- unique(unlist(destinos[row, 'nColPT'], use.
  names = F))
124 for (b in xdest) {
125 linha <- data.table(id=xid, bairro=b)
126 ndest <- rbind(ndest, linha)
127 }
128 }
129 tmp9 <- left_join(ndest, tmp7t, by=c('id', 'bairro'))
130

```



```
131 # Unindo origens e destinos
132 tmp10 <- left_join(tmp8, tmp9, by='id')
133
134 # Selecionando as colunas "informações" que interessam
135 tmp11 <- select(tmp10, -c('data.x', 'dia.x': 'local.x',
  'local.y'))
136 colnames(tmp11) = c("id", "origem", "Lat_0", "Long_0", "
  destino", "data_D", "Lat_D", "Long_D",
137 "dia_D", "mes_D", "ano_D", "dsm_D", "hora_D")
138
139 # Salvando as viagens em um arquivo do tipo CSV
140 write.csv2(tmp11, paste(pasta, "NovoInstagram.csv", sep=
  ""), row.names = FALSE)
141
142 # FIM
```