**Título**

**RESUMO**

**Palavras-chave**: *Mobilidade urbana, Simulação de dados, Big Data Analytics, Estatística Espacial.*

# 

# **1. INTRODUÇÃO**

**1.1. Problema da Pesquisa**

**1.2. Justificativas**

**2. REVISÃO DE LITERATURA**

**2.1. Mobilidade Urbana e Acessibilidade**

**Metrô. (2008). Pesquisa Origem-Destino 2007. Secretaria de Transportes Metropolitanos. Retrieved from http://www.metro.sp.gov.br/ pesquisa-od/arquivos/OD\_2007\_Sumario\_de\_Dados.pdf**

* Definições legais de mobilidade e acessibilidade no Brasil
  + - PNMU
* Definições e discussões na literatura de Mobilidade e acessibilidade
  + Mobilidade
    - SILVEIRA, M. R., COCCO, R. G. Transporte público, mobilidade e planejamento urbano: contradições essenciais. Estudos avançados, São Paulo, v. 27, n. 79, p. 41-53, 2013. Disponível em <http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S010340142013000300004&lng=en&nrm=iso>. Acesso em 1 Junho de 2018.
    - GAKENHEIMER, R. Urban mobility in the developing world. Transportation Research Part A: Policy and Practice, 33(7-8), 671-689, 1999.
    - Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A Tale of many cities: Universal patterns in human urban mobility. PloS One, 7(9). doi:10.1371/annotation/ca85bf7a-7922-47d5-8bfbbcdf25af8c72
  + Acessibilidade
    - HÄGERSTRAAND, T. What about people in regional science? Papers in regional science, v. 24, n. 1, p. 7-24, 1970.
    - Kwan, M. (1998). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. Geographical Analysis, 30(3), 191-216. doi:10.1111/j.1538-4632.1998. tb00396.x
    - Litman, T. (2003). Measuring transportation: Traffic mobility and accessibility. Victoria Transport Policy Institute. ITE Journal, 73(10), 28-32.
    - Lessa, D. A., Lobo, C., & Cardoso, L. (2019). Accessibility and urban mobility by bus in Belo Horizonte / Minas Gerais – Brazil. Journal of Transport Geography, 77(May 2018), 1-10. doi:10.1016/j. jtrangeo.2019.04.004
    - Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. Journal of Transport Geography, 25, 141–153. https://doi. org/10.1016/j.jtrangeo.2012.03.016
    - WANG, M., & MU, L. Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA. Computers, Environment and Urban Systems, 67, 169-175, 2018.
    - Hughes, R., & MacKenzie, D. (2016). Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators. Journal of Transport Geography, 56, 36-44. doi:10.1016/j. jtrangeo.2016.08.014
    - TRIBBY, C. P., ZANDBERGEN, P. A. High-resolution spatio-temporal modeling of public transit accessibility. Applied Geography, 34, 345-355, 2012.

**2.2. Indicadores de acessibilidade**

* **Diferentes implementações** 
  + - Páez, A., Scott, D. M., & Morency, C. (2012). Measuring accessibility: Positive and normative implementations of various accessibility indicators. Journal of Transport Geography, 25, 141–153. https://doi. org/10.1016/j.jtrangeo.2012.03.016
    - HÄGERSTRAAND, T. What about people in regional science? Papers in regional science, v. 24, n. 1, p. 7-24, 1970.
    - Kwan, M. (1998). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. Geographical Analysis, 30(3), 191-216. doi:10.1111/j.1538-4632.1998. tb00396.x
    - TRIBBY, C. P., ZANDBERGEN, P. A. High-resolution spatio-temporal modeling of public transit accessibility. Applied Geography, 34, 345-355, 2012.
* Acessibilidade, segregação e indicadores sociais
  + HEILMANN, K. (2021). Regional Science and Urban Economics Transit access and neighborhood segregation. Evidence from the Dallas light rail system. Regional Science and Urban Economics, 73, 237–250, 2018. https://doi.org/10.1016/j.regsciurbeco.2018.10.007 JÚNIOR
  + **TORRES, H. D. G., MARQUES, E., FERREIRA, M. P., BITAR, S. Pobreza e espaço: padrões de segregação em São Paulo. Estudos avançados, 17(47), 97-128, 2003.**
  + **TORRES, H. D. G., & OLIVEIRA, G. C. D. Primary education and residential segregation in the Municipality of São Paulo: a study using geographic information systems. In International Seminar on Segregation in the City, pp. 26-28, Julho de 2001.**
  + **Mestrado da eliana**
  + **Morandi, E., Ribeiro, R., Hernandes, E., Camara, B., Spinola, L., & Francisco, E. D. R. (2016). Análise Geoespacial da Relação entre Transporte Público sobre Trilhos, Renda e Tempo Médio de Deslocamento. Paper presented on XL EnANPAD, Salvador, BA. Retrieved from http://www.anpad.org.br/~anpad/eventos. php?cod\_evento=1&cod\_edicao\_subsecao=1302&cod\_evento\_ edicao=83&cod\_edicao\_trabalho=21023**
  + **Preston, V., & McLafferty, S. (1999). Spatial mismatch research in the 1990s: Progress and potential. Papers in Regional Science, 78(4), 387-402. doi:10.1111/j.1435-5597.1999.tb00752.x**
  + André Leite Guerra¹, Heloisa Maria Barbosa², Leise Kelli de Oliveira³ - Levantam diferentes estudos sobre o uso de bancos de dados gerados por sistemas de bilhetagem eletrônica para simular matrizes de origem e destino das viagens de transporte público. Além disso, produzem um estudo dessa natureza para a cidade de Maceió, encontando resultados positivos, mas com a determinação das limitações da abordagem – algums suposições são necessárias para completar dados de viagens “incompletas” (só com a entrada no sistem) e para definir ponsots de entrada e saída nos sistema de ônibus; os resultados finais também não tem uma forma de validação que garanta a incerteza presente na matriz final – o que diminui a confiança da estratégia.
  + Renato Oliveira Arbex1, Claudio Barbieri da Cunha2 – propoê uma metodologia de estimação de uma matriz OD para as estações do sistema metroferroviário utilizando dados do sitema de bilhetagem- que só valida a entrada no sistema e não a saída - complementado com registros de entradas no sistema de ônibus, para identificar saídas. A partir dessa matriz os autores calculam os padrões de carregamento das várias rotas do sistema metroferroviário durante um dia comum de viagens. São feitas algumas suposições sobre o comportamento dos usuários quanto à demora das viagens através de médias conhecidas em outros estudos. Foi também usado o caminho de menor tempo para ligar as origem e destinos das viagens. A proposta permite acompanhar a distribuição da demanda de transporte metroferroviário exigida do sistema.

**2.3. *Big Data* e Estatísticas Oficiais**

* **Big Data como nova estratégia analítica**
  + **MCAFEE, A., BRYNJOLFSSON, E., DAVENPORT, T. H., PATIL, D. J., BARTON, D. Big data: the management revolution. Harvard business review, 90(10), 60-68, 2012.**
  + **GANDOMI, A., HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, v. 35, n. 2, p. 137-144, 2015**
  + **Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. Big data research, 2(4), 166-186. doi:10.1016/j. bdr.2015.01.001**
  + **Letouzé, E., & Jütting, J. (2015). Official statistics, Big Data and human development: towards a new conceptual and operational approach. Data Pop Alliance White Paper Series. Retrieved from https://paris21. org/sites/default/files/WPS\_OfficialStatistics\_June2015.pdf**
  + **LEE, J. G., KANG, M. Geospatial big data: challenges and opportunities. Big Data Research, 2(2), 74-81, 2015**
* **Discussão da relação de Big Data com Estatísticas oficiais**
  + **Letouzé, E., & Jütting, J. (2015). Official statistics, Big Data and human development: towards a new conceptual and operational approach. Data Pop Alliance White Paper Series. Retrieved from https://paris21. org/sites/default/files/WPS\_OfficialStatistics\_June2015.pdf**
  + **DARGENT, E., LOTTA, G. , MEJÍA, J. A., MONCADA, G. A quem importa saber?: a economia política da capacidade estatística na América Latina, 2018**
  + **Kim, B. G., Trimi, S., & Chung, J. (2014). Big-Data applications in the government sector. Communications of the ACM, 57(3), 78-85. doi:10.1145/2500873**
* **As incertezas do Big Data**
  + **Kwan, M. P. (2016). Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. Annals of the American Association of Geographers, 106(2), 274-282. doi:10.10 80/00045608.2015.1117937**

# **3. MÉTODO**

O processo de comparação dos dados simulados seguiu três etapas: a simulação e organização dos dados derivados da API, onde está presente o potencial de Big Data do processo; a organização dos microdados da Pesquisa OD e comparação com os dados simulados; e a análise propriamente da compração entre as medidas de tempo simuladas e as medidas presentes na Pesquisa OD.

**3.1. Simulação de viagens**

A simulação foi feita em duas etapas; Primeiro foi gerado um banco de endereços no município de São Paulo utilizando o software QGis; a partir dessa base de endereços foi feita a simulação das viagens propriamente ditas utilizando a API de roteamento do Google. A escolha dessa ferramenta levou em consideração a possibilidade de simular viagens considerando o efeito do tráfego em tempo real das viagens, fator que foi considerado um diferencial da estratégia adotada. Porém isso implicou em um número limitado de requisições de viagens, por questões de custo; essa limitação levou a uma série de escolhas para reduzir o número de viagens “perdidas” na simulação, quando as coordenadas usadas na API não correspondiam ou não podiam ser aproximadas a endereços válidos (como no caso de coordenadas nas represas de São Paulo). Outra escolha da simulação, de distribuir as viagens simuladas ao longo dos dias úteis (segunda-feira à sexta-feira, das 5h às 21h) levou à opção por usar computação em nuvem para a simulação levantando a preocupação de reduzir o esforço computacional evitar problemas relacionados ao desempenho. Essas limitações definiram o processo de definição de endereços. Procurou-se sortear endereços em regiões mais densamente povoadas para evitar possíveis perdas, seguindo um modelo de densidade de probabilidade da população. Ao mesmo tempo, para reduzir o esforço computacional, foi montada uma base de coordenadas *offline*, que foi usada para sortear os endereços das viagens. O processo de geração da base foi feito nas seguintes etapas: 1) O mapa do município de São Paulo (em formato *shapefile*) foi dividido por uma grade com quadrículas de 500 metros de lado; 2) Foram calculadas as populações de cada quadrícula com dados do Censo Demográfico de 2010 e retiradas as quadrículas com população igual a zero; 3) As quadrículas foram divididas em quintis de densidade populacional, e sorteados aleatoriamente pontos geográficos dentro de cada quadrícula, de acordo com o quintil: 5 pontos para o quintil mais populoso, 4 para o 2º quintil, 3 para o 3º, 2 para o 2º e 1 ponto para o quintil menos populoso. 4) O conjunto de pontos resultante foi usado como base para o sorteio dos endereços de origem e destino. Essa primeira etapa foi realizada utilizando bases cartográficas abertas do município de São Paulo e o software aberto QGIS. A simulação das viagens foi feita a partir de um programa desenvolvido em Python, executado no serviço de computação em nuvem da Google. A estrutura do programa seguiu a arquitetura apresentada na Figura 1. Foi usada uma ferramenta de agendamento (Google Scheduler) de ativação ligada a uma máquina virtual no ambiente em nuvem da Google, para que a chamada ao serviço de viagem do Google Maps fosse realizada nos dias úteis da semana, a cada hora cheia, das cinco da manhã até às nove da noite. O programa seguiu as etapas descritas na Figura 1.

* Abrir uma conexão com o banco de dados SQL da nuvem da Google
* Carregar o banco de coordenadas
* Sortear dez coordenadas de origem e dez coordenadas de destino
* Chamar a API Distance Matrix com as dez origens e os dez destinos, para viagens de transporte público
* Processar os resultados devolvidos pela API e armazenar em um vetor auxiliar
* Chamar novamente a API Distance Matrix com as mesmas dez origens e os dez destinos, para viagens de transporte privado
* Processar os resultados devolvidos pela API e anexar ao vetor auxiliar
* Submeter o vetor auxiliar à função que insere os dados no Banco de Dados hospedado na nuvem.

**Figura 1.**Etapas e Estrutura funcional do programa elaborado para simular os dados  
Fonte: Os autores

Cada chamada da API Distance Matrix retornou uma lista com duzentas viagens com dados de horário e dia da semana, coordenadas da origem e do destino, endereços da origem e do destino da viagem, duração, distância e tarifa da viagem. Para cada par origem-destino houve registro de viagem de carro e de transporte público. O período de simulação foi entre os dias 11 de fevereiro de 2019 a 5 de junho de 2019. O total de viagens armazenadas no banco de dados nesse período foi de 257.400 viagens, sendo 253.450 viagens válidas – 128.700 (100% de aproveitamento) das viagens de carro e 126.725 das viagens transporte público (98,47% de aproveitamento).

**3.2. Organização dos microdados da Pesquisa Origem-Destino e Comparação**

Os microdados da Pesquisa OD apresentam entradas correspondentes a cada viagem realizada por indivíduo entrevistado (caso o indivíduo não tenha realizado nenhuma viagem, a entrada só corresponde aos dados socioeconômicos do entrevistado). Cada linha do banco registra sobre o indivíduo, de forma anônima, sua identidade, a família na qual ele foi registrado na pesquisa, o domicilio em que ele reside, características físicas do domicílio, características socioeconômicas e ocupacionais do indivíduo e da família; sobre a viagem são registrados o propósito da viagem, o modal principal de deslocamento, o tempo da viagem e dos trechos à pé necessários para acessar os modais usados; as baldeações realizadas, as coordenadas,zonas e municípios da origem e do destino da viagem, a hora de saída e chegada, entre outras (CITAÇÂO).

Para o presente trabalho foram filtradas informações relativas somente às viagens – modais principais, tempo de duração (da viagem e dos trechos a pé), horário de saída e coordenadas das origens e destinos. A partir dos modais principais, foram criados dois bancos de dados, um relativo ao transporte público (Metrô, Trem, Monotrilho, Ônibus/micro-ônibus/peruas municipais e metropolitanos) e transporte privado (Automóvel – passageiro ou motorista, táxi convencional e táxi não convencional). A opção dessa agregação foi feita em função da classificação modal que foi possível a partir dos dados da simulação das viagens.

O passo seguinte foi a espacialização dos bancos de dados, tanto o simulado como os bancos privado e público da OD. Utilizando o software aberto RStudio, foram utilizadas as coordenadas de origem e destino nos bancos para criar dois *Shapefiles* para cada banco, um a partir dos pontos de origem das viagens e outro a partir dos destinos, com os mesmos dados associados. Para a comparação do banco simulado com os bancos OD, foi criado um algoritmo, repetido para cada viagem dos Bancos OD.

Dada uma viagem OD, foram selecionadas todas as viagens simuladas com a origem dentro de um *buffer* de 1500m (escolhido por incluir 95% dos trajetos a pé para acessar os transportes); dessas viagens foi selecionado ainda o subconjunto de viagens simuladas em que os pontos de destino também se encontrassem dentro de um *buffer* de 1500m do ponto de destino da viagem OD. A partir desse conjunto de viagens simuladas foi calculada a média da duração simulada (considerando se a viagem comparada era de transporte público ou privado), que por sua vez foi subtraída da soma da duaração da viagem OD, do tempo andando da origem e do tempo até o destino da viagem OD, convertidos em segundos. Esse algoritmo foi repetido para todas as viagens OD e a diferença calculada foi registrada para cada viagem. Uma variante desse algoritmo foi realizada também, adicionando mais uma restrição para a seleção das viagens: além dos *buffers* espaciais, foi utilizado um *buffer* temporal, filtrando as viagens simulada que começassem duas horas antes ou depois da viagem OD.

**3.3. Análise dos dados**

A partir da diferença calculada entre a duração das viagens OD e das médias das viagens simuladas, foram realizadas medidas descritivas para analisar o comportamento da variável. Foram elaboradas dispersões entre a diferença e as outras variáveis dos bancos para verificar possíveis dependências nos dados.

Foi feita uma análise espacial utilizando o software aberto GEODA para verificar a existência de dependência espacial; para isso foram criadas vizinhanças entre os pontos de origem das viagens considerando 5, 20 e 40 vizinhos, e considerando *buffers* de 500m e 1500m. A partir das análises do I de Moran, não foi identificada nenhuma dependência espacial.

Por último foram realizados testes T de diferença de médias para a diferença calculada entre as bases, tanto para o caso com restrição de tempo como para o sem. Foram comparados os resultados nesses dois casos, com resultados significativos.

Para todas as análises foram comparados os resultados entre os transportes públicos e privados, como mostrado na seção seguinte.

# **4. RESULTADOS E ANÁLISE**

# **5. DISCUSSÃO FINAL E PRINCIPAIS CONCLUSÕES**

# **REFERÊNCIAS BIBLIOGRÁFICAS**