**Programa Institucional de Bolsas de Iniciação Científica (PIBIC)**

**Simulação de dados de mobilidade como estratégia para produção de análises sobre o transporte público**

Aluno: Rodolfo Oliveira Lorenzo

Orientador: Eduardo de Rezende Francisco

Campo de estudo: Administração Pública e Estatística

Pergunta de Pesquisa:

Usando ferramentas e algoritmos de Estatística Espacial, é possível utilizar a simulação de tempos de viagem para gerar informações relevantes sobre as viagens no município de São Paulo?

Objetivo Geral

Os dados de mobilidade são um insumo importante para o planejamento da expansão da infraestrutura urbana, tanto é assim que a empresa do Metropolitano de São Paulo é responsável por realizar a cada dez anos a Pesquisa de Origem e Destino (Pesquisa OD) sobre mobilidade na grande São Paulo. A pesquisa é referência para trabalhos acadêmicos e projetos de infraestrutura, mas devido a sua dimensão e custo, acontece a cada dez, apresentando um problema de periodicidade, mesmo sendo intercalada com uma pesquisa menor para atualiza-la, mas menos desagregada, no intervalo entre cada duas edições. Assim o presente trabalho buscará validar métodos de gerar duas das variáveis da Pesquisa OD, o tempo de viagem e a distância das viagens para diferentes modais, de forma que essa grandeza possa ser usada em futuras pesquisas, sendo geradas quando necessário.

Atividades realizadas

| **Atividades previstas** | | **Estado** | |
| --- | --- | --- | --- |
| Pesquisar referencias | | Em realização | |
| Programa de cálculo de tempos de viagem | Sorteio de pontos | | Feito |
| Calculo de viagens | | A fazer |
| Coleta de dados OD | | Em espera | |
| Levantamento de variaveis | | A realizar | |
| Modelagem de dados | | A realizar | |
| Análise de resultados | | A realizar | |

Em relação ao cronogram inicial, o projeto esta atrasado. Isso se deu principalmente em razão da maior complexidade encontrada para a formulação do programa de cálculo de viagens do que o esperado. Para que o programa consiga gerar dados úteis para a análise, ele precisa funcionar com uma peridiocidade definida e prevista em suas chamadas; sua execução de ser ininterrupta e deve atinjir períodos relevantes para a análise, como os dias e as horas em que ocorre a mobilidade urbana. A implementação dessas funcionalidades demonstoru ser mais complexa que o esperado. A idéia é utilizar feramentas de computação em nuvem disponibilizadas em caráter gratito por empresas de tecnologia para que o programa apresente essas caracterísiticas.

Em relação a levantamento teórico, conceitos relevantes à pesquisa foram encontrados em ainda precisam de maior referenciamento. Por essa razão, essa fase de trabalho també precisará de mais tempo de execução para que seja completada a contento.

As variáveis da pesquisa OD ainda não puderam ser analisadas em razão de não terem sido publicadas de forma desagregada até o momento. Caso isso não seja feito até uma data que permita a análise por esse projeto, serão usdas outras edições da pesquisa.

Escopo do projeto e referencial teórico

O referencial teórico está dividido nas tematicas relevante ao trabalho.

A questão da mobilidade em São Paulo

A mobilidade na metrópole de São Paulo é resultado de seu processo de urbanização e reflete as vicissitudes do mesmo. Durante o século XX o desenvolvimento das cidades brasileiras seguiu um padrão semelhante de crescimento intenso e periferização precária, gerando uma ocupação segregada do espaço urbano (Maricato, 2003; Rolnik e Klink, 2011). São Paulo, como principal expoente do crescimento urbano do período, não fugiu desse padrão. Um intenso processo migratório acompanhado de uma rápida industrialização, cujo processo de capitalização drenava os recursos disponíveis, levou a formação de periferias extremamente vulneráveis, com péssimas condições de habitabilidade, além de distâncias consideravelmente grande ao centro (Kowarick, 1979). Mesmo considerando que a condição das periferias do município não foi uniformemente constituída, e que houve intervenções do Estado relativas à oferta de infraestrutura e serviços desde dos anos 70, ainda sim nos anos 2000 os indicadores das periferias apresentavam números consideravelmente piores que das áreas centrais da cidade, com exceção particular das regiões centrais ocupadas por favelas e cortiços (Torres e Oliveira, 2001; Torres *et al.*, 2003). O crescimento da mancha urbana de São Paulo, em seu processo de conurbação, levou a lógica da periferização para os municípios vizinhos, seguindo tendência já apontada no fim dos anos 70 (Kowarick, 1979), quando os indicadores sociais das periferias do município de SP, que eram muito piores que as áreas centrais, passaram a se estender para a periferia estendida, nos municípios vizinhos Outro fator importante na relação centro-periferia que deriva também dessa formação urbana é a concentração de trabalhos disponíveis no Município de São Paulo. Apesar de em muito dos municípios vizinhos terem desenvolvido importantes economias, inclusive industriais (como no grande ABC), o censo de 2000 mostra que dos quase um milhão de habitantes da RMSP que trabalhavam ou estudavam fora do município de origem, mais da metade se dirigia para o município de São Paulo (Aranha, 2005) - tendência centro-periferia metropolitana que se reproduz na escala municipal.

Dentro dessa demanda intensa por mobilidade, as decisões tomadas em relação à questão agravaram o cenário. Por muito tempo foi priorizada a mobilidade viária em detrimento dos trens e do metrô, com grandes projetos de expansão viária e anéis de circulação, e dentro dessa foi incentivado o uso de transporte individual, em razão de incentivos à indústria automobilística a nível nacional, em detrimento do coletivo (Júnior, 2011; Gakenheimer, 1999; Silveira e Cocco, 2013; Wilheim, 2013; Scaringella, 2001), o que gerou uma infraestrutura viária incapaz de acompanhar as taxas de motorização do Brasil, além de um sistema de transporte público dependente de um empresariado ligado aos ônibus (Silveira e Cocco, 2013). Essa situação levou ao agravamento das condições de mobilidade para os moradores das periferias paulistanas que, dependendo tanto de carros como de transportes público, sofrem com viagens longas, congestionamento e saturação dos meios coletivos. Essa condição não é particularidade do Brasil, ou de São Paulo: cidades que passaram por intensos processo de urbanização associados a motorização apresentam grandes dificuldades para manter sua infraestrutura de transportes em compasso com a demanda (Gakenheimer, 1999).

A produção de estatísticas oficiais e o Big Data

A capacidade de produção de estatísticas oficiais confiáveis e periódicas é um fator essencial para a capacidade de um país tomar decisões racionais em relação ao futuro, baseada em evidências capazes de indicar algo da realidade (Dargent *et al.*, 2018). Em termos do Estado essa capacidade atende necessidades tanto para o desenvolvimento de novas políticas públicas como para o monitoramento e avaliação das existentes – em relação ao Brasil, a contabilidade populacional e a previsão de sua evolução são dados importantes para o repasse de recursos federais para os municípios. Para a sociedade civil e para o mercado, a produção de dados confiáveis permite que se realizem pesquisas relevantes aos diversos atores sociais e planejamanto futuro em relação a evolução dos indicadores derivados desses dados. Ainda, para vários países e órgãos multilaterais, a participação em programas de ajuda financeira, ou mesmo parcerias dentro do setor privado, exigem a presença de indicadores sociais e economicos confiáveis. De fato, tanto a necessidade interna do Estado como e demanda de atores externos ao Estado, ou externos ao país, são identificados como fatores de economia política que explicam o desenvolvimento dessa capacidade dentro do Estado (Dargent *et al.*, 2018).

Ao mesmo tempo Letouzé e Jütting (2014) discutem uma “desilusão estatísitica”: há um descontentamento com a capacidade das burocracias estatais em produzir estatísitcas confiáveis e relevantes – desde de modelos tradicionais que não conseguem acompanhar períodos voláteis até medidas que são consideradas insuficientes para o que se propoê, como o PIB para medir bem estar. Ainda, em países pobres e em desenvolvimento essa desilusão está associada a baixa capcaidade atual, que gera situações como a de Gana, em que a adoção de uma metodolgia mais nova de calculo de PIB indicou um crescimento de 60% desse[[1]](#footnote-2). A dificuldade desses países de construir essa capcidade passa pela falta de recursos financeiros, a baixa capacitação técnica do serviço público (causa e conseqüencia de uma fuga de cérebros para o setor privado), intervenções políticas na produção de dados, entre outras (Letouzé e Jutting, 2014). Além disso, nesse cenário de fragilidade e coloca a crescente produção de dados e as novas formas de análises estatísitcas baseadas nesse “movimento”: Big Data.

Uma das primeiras definições de Big Data esta relacionada as caracterśiticas dos dados englobados pela definição. O aumento da produção, capacidade de armazenamento e processamento de dados gerou a potencialide de aplicações analíticas que, se não apresentam necessariamente métodos inovadores em termos estatísticos, conta com inovações computacionais e três grande conceitos definidores em relação ao dados envolvidos: Volume, Velocidade e Variedade (McAfee *et al.,* 2012; Gandomi e Haider, 2015). De acordo com essa definição, o que caracteriza Big Data é não só o volume dos dados envolvidos, mas também a velocidade de produção de dados, com aplicaçẽos analisando dados produzidos em tempo real, e a variedade de formatos, com o uso de dados estruturados e não estruturados . Ainda nessa direção existem definições que inclem Veracidade (em relação a dados como o estado socioemocional de usuários de redes sociais, que mesmo tendo valor apresentam um graud e incerteza quanto ao seu conteúdo), Variabilidade e Complexidade (variabilidade em relação aos rtimos do fluxo de dados e complexidade em relação ao uso de diversas fontes para os dados, o que exige trabalho para agrega-los) e Valor (Em relaçao ao baixo valor de um dado singular em comparação com o valor que o grande agregado possui)(Gandomi e Haider, 2015). Existem outras definições de Biga Data, que partem de outros pressupostos. Letouzé e Jütting (2014) o definem a partir de caracteŕisticas “sociológicas. Os três conceitos definidores de Big Data seriam a natureza dos dados (não o volume), que são gerados como rastros de atividade humana dentro da rede (como o comportamento em redes sociais) – “Crumbs” ou migalhas; as técnicas e a intenção envolvida na geração de “insights” a partir desses dados, que envolvem capacidades avançadas de armazenamento e computaçã e métodos e ferramentas quantitativos e computacionais avançados - “Capacities”; esses dados e essa técnicas são utilizados por comunidades específicas relacionadas ao desenvolvimento dessas aplicações, tanto dentro da comunidade de softwares abertos como dentro dos setor privado e de inteligencia - “Comunities” - os três C’s. Outras definições partem ainda de critérios voltados a implementação de sistemas, com a classificação de arquiteturas de Big Data (Pääkkönen e Pakkala, 2015).

A relação entre as estatísticas oficiais e o Big Data pode ser vista como representativa do conflito sobre a capacidade do Estado de fornecer dados ágeis e úteis. Por um lado o Big Data é capaz de produzir informações a partir de dados produzidos em tempo real, coletados automaticamente de diversas fontes. É possível, a partir dessa capacidade, tentar reproduzir os indicadores oficiais já existentes, ou outros, mais granulares e inteligentes. Letouzé e Jütting (2014) argumentam, porém, que a responsabilidade das agências oficiais, ao produzir os dados oficiais, não é só de gerar informações úteis: Elas tem a função de produzir conhecimento sobre a sociedade, e além disso de constituir um espaço deliberativo sobre o que merece ser medido na sociedade. Nesse sentido, pensando no movimento de Big Data como um importante vetor de mudança na sociedade moderna, é interessante que haja movimentos de integração entre as estatísticas oficiais e essas novas técnicas de análise.

De particular interesse para o presente trabalho, a produção de dados georreferenciados relativos a mobilidade é essencial para captar a distribuição da mobilidade no tecido urbano. Dentro dos meios de Big Data, os dados gerados pela utilização dos celulares – ainda mais no contexto em que volume da rede móvel supera o volume de rede fixa (Lee & Kang, 2015) - já fornece um enorme volume de dados georreferenciados e, dependendo do uso de aplicativos, com informações sobre os meios de transporte. Essa produção massiva de dados permite inclusive o uso desses dados para análises em tempo real, como os serviços de mapas para calcular rotas. Também pelo lado das estatísitcas oficiais a produção de dados georreferenciados para entender os problemas urbanos, inclusive de mobilidade, é corrente e importante para embasar a adoção de políticas públicas específicas para cada localidade. A compreensão da dimensão geográfica dos problemas e da distribuição da infraestrutura presente e dos serviços ajudam a diagnosticar ineficiências e priorizar esforços, além de fornecer uma visão sistêmica dos indicadores sociais. Essa visão pode ajudar a escolher combinações de formas diferentes de intervenção pública (Torres *et al.*, 2003, Torres e Oliveira, 2001). Mas a produção desses dados através de pesquisas empíricas de validade estatística, como a Pesquisa OD (METRO, 2008), tende a ser bem custosa. O acesso a dados derivados dos novos aplicativos sociais que usam a localização podem permitir o acesso a informações de mobilidade de maneira muito mais barata, ainda que contendo algum grau de viés - esses dados podem fornecer informações valiosas sobre os padrões de mobilidade e acessibilidade das cidades (Noulas, Scellato, Lambiotte, Pontil, Mascolo, 2012; Wang e Mu, 2018).

Mobilidade e Acessibilidade

Em relação à mobilidade, a compreensão das formas de usos de diferentes modais em cada região podem ajudar a associar os padrões de mobilidade a certos grupos sociais, permitindo pensar em políticas voltadas para equilibrar os usos do espaço público para melhorar a mobilidade de quem mais precisa. Em São Paulo, estudos nessa direção identificam a dependência mais acentuada dos moradores periféricos de modais coletivos em relação aos individuais, mas também identificam uma expressiva periferia motorizada, que demanda espaço urbano para sua mobilidade (Requena, 2015). Há a associação entre os tempos médios de viagem e a acessibilidade a rede de transportes rápidos (trem e metrô) nos distritos de São Paulo, e essas por sua vez têm associação com as rendas médias dos distritos, o que contribui para uma distribuição desigual da mobilidade (Morandi *et al.*, 2013).

Mas entender a mobilidade urbana, apesar de sua importância, não engloba toda a experiência de acesso a cidade. A informação de como os indivíduos se locomovem na cidade não nos informa se eles conseguem acessar as oportunidade que a cidade pode oferecer; um conceito mais amplo, capaz de refletir o acesso dos indivíduos à cidade é a acessibilidade (Litman, 2003). A mobilidade, de acordo com a definição de Litman, é um meio para que os indivíduos cheguem aos seus destinos. Assim, para o estudo da acessibilidade, o que interessa em relação a mobilidade é o custo - tempo, dinheiro, desconforto ou risco - que ela implica aos indivíduos, e esse custo é um dos componentes das medidas de acessibilidade; o outro componente é a qualidade e a quantidade de oportunidades e sua distribuição no tecido urbano (Paéz, Scott e Morency, 2012).

As medidas de acessibilidade podem ser elaboradas baseadas nos indivíduos, associando a ele o valor da medida, ou baseadas nos lugares, em que a acessibilidade é um atributo do lugar; ao mesmo tempo, as medidas podem ser centradas no local da origem das viagens potenciais ou no local de destino das viagens. Também, os dois componentes das medidas, o custo de transporte e a distribuiçao de oportunidades, podem ser abordados de forma normativa ou positiva. A abordagem positiva consite em considera o que acontece, tanto em termos da mobilidade como da distribuição de oportunidades. A abordagem normativa considera o que deveria acontecer (em termos de mobilidade, qual é custo que deveria ser aceitável para o indivíduo) e em geral não se utiliza na distribuição das oportunidades (Paéz, Scott e Morency, 2012). Em relação aos tipos de indicadores de acessibilidade, a literatura abordada até o momento aponta quatro grupos: os indicadores “gravitacionais”, os indicadores cumulativos, os indicadores baseados em utilidade e inidicadores de espaço-tempo. Os indicadores gravitacionais,os cumulativos e os de espaço tempo são instâncias particulares da seguinte fórmula (Paéz, Scott e Morency, 2012; Kwan, 1998):

A medida de Acessibilidade A é dada para a origem i e as oportunidades k para o indivíduo p em função do numero de oportunidades W no local j dado dentro de uma função de atratividade g, multiplicada por uma função de impedência f, que é um kernel em volta da origem i dado em função do custo de viagem c do local i para o j para a população de p.

Para os indicadores gravitacionais, a função g é uma função de atratividade do local j que é dada em função da oportunidades k presentes. A função de impedância costuma ser uma função que varia de algum valor positivo na origem a 0 no infinito – por exemplo, uma exponencial negativa, ou uma potência invertida, ou uma gaussiana modificada (Kwan, 1998). Já para os indicadores cumulativos, a função f é um inequação simples em que seu valor é 1, se c está dentro de certo limite pré definido, o 0 se c está fora – o valor do indicador se refere ao numero de oportnidades que estão dentro doa raio de custo definido. Para os indicadores de espaço tempo, o custo c pode ser usado como uma região dentro de uma rede correspondente à área de caminho potencial (PPA) (Hägerstraand, 1970; Kwan, 1998), que reflete a área que o indivíduo pe capaz de acessar dados os seus constrangimentos diários. Enquanto as duas primeiras medidas são baseadas em lugares, essa última e feita em relação aos indivíduos. As medidas de utilidade são baseadas no termo “log-sum” de “modelos discretos de escolha aplicados à análise se escolha de destino” (Paéz, Scott e Morency, 2012).

Alguns problemas dos indicadores relativos à lugares, como os cumulativos e de gravidade, é que eles ignoram as especificidades da mobilidade de indivíduos nas áreas analisadas. Por exemplo, casos específicos em que as mulheres consistentemente mostram padrões diferentes de acessibilidade, mesmo morando nas mesmas regiões, ou mesmo mesmas casa, que homens (Kwan, 1998; Paéz, Scott e Morency, 2012). Ao mesmo tempo, o uso de uma referência de origem impede que os indicadores deem conta de comportamentos de mobilidade diferentes do padrão casa-trabalho. Além disso, como a implementação costuma ser feita a partir de dados agragados em métodos zonais, existem problemas de escolha de limites e possíveis falácias ecológicas (Kwan, 1998). Os indicadores de espaço-tempo, apesar de contornarem alguns desses problemas, já que são baseados nos indivíduos e consideram os diferentes tipos de comportamento, apresentam uma implementação computacionalmente muito mais complexa e custosa, além de entregarem resultados que são menos capazes de caracterizar os lugares (Kwan, 1998).

Resultados parciais

Sorteio de pontos no município de São Paulo

Foi aproveitado uma base de dados de pontos de endereços em São Paulo montada para um trabalho semelhante. A base foi construída pensando na análise a partir dos distritos de São Paulo, e por isso tem uma preocupação de garantir que todos os distritos tenha pelo menos algum ponto sorteado aleatóriamente. Para isso, a cidade foi dividida em 1720 quadrículas de 1km² e dentro de cada quadricula foi sorteado um ponto aleatóriamente. A partir dessa base de pontos serão calculadas as viagens.

Programa de calculos de viagens

O programa será feito usando conceitos de cloud computing. Estão sendo estudada a possibilidade de utilizar as ferramentas disponibilizadas pelo Google Cloud de forma gratuita para montar o programa.

Referências Bibliográficas

ARANHA, V. Mobilidade pendular na metrópole paulista. *São Paulo em perspectiva*, v. 19, n. 4, p. 96-109, 2005.

BADDELEY, A., TURNER, R. Spatstat: an R package for analyzing spatial point patterns. *Journal of statistical software*, *12*(6), 1-42, 2005.

BADDELEY, A. Analysing spatial point patterns in R. Technical report, CSIRO, 2010. Version 4. Fevereiro de 2008. URL https://research. csiro. au/software/r-workshop-notes.

CÂMARA, G., MONTEIRO, A. M., FUCKS, S. D., CARVALHO, M. S. Spatial analysis and GIS: a primer. *National Institute for Space Research. Brasil,* 2004.

CIA. DO METROPOLITANO DE SÃO PAULO. Pesquisa Origem-Destino 2007. São Paulo: Secretaria de Transportes Metropolitanos, 2008

DARGENT, E., LOTTA, G. , MEJÍA, J. A., MONCADA, G. A quem importa saber?: a economia política da capacidade estatística na América Latina, 2018

FRANCISCO, E. R. Indicadores de renda baseados em consumo de energia elétrica: Abordagens domiciliar e regional na perspectiva da estatística espacial. 2010. 381 f. Tese (Doutorado em Administração de Empresas) - Escola de Administração de Empresas de São Paulo, Fundação Getulio Vargas, São Paulo, 2010.

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.

GAKENHEIMER, R. Urban mobility in the developing world. *Transportation Research Part A: Policy and Practice*, *33*(7-8), 671-689, 1999.

HÄGERSTRAAND, T. What about people in regional science?. Papers in regional science, v. 24, n. 1, p. 7-24, 1970.

JÚNIOR, J. A. O. Direito à mobilidade urbana: a construção de um direito social1. *Revista dos Transportes Públicos-ANTP-Ano*, *33*, 1º, 2011.

KWAN, M. P. Space‐time and integral measures of individual accessibility: a comparative analysis using a point‐based framework. Geographical analysis, v. 30, n. 3, p. 191-216, 1998.

LEE, J. G., KANG, M. Geospatial big data: challenges and opportunities. *Big Data Research*, *2*(2), 74-81, 2015.

LETOUZÉ, E., JÜTTING, J. Official statistics, big data and human development: towards a new conceptual and operational approach. Data Pop Alliance and PARIS21, 2014.

Litman, T. Measuring Transportation: Traffic Mobility and Accessibility. Victoria Transport Policy Institute, 2003.

MARICATO, E. Metrópole, legislação e desigualdade. *Estudos avançados*, *17*(48), 151-166, 2003.

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.

NOULAS, A., SCELLATO, S., LAMBIOTTE, R., PONTIL, M., MASCOLO, C. A tale of many cities: universal patterns in human urban mobility. *PloS one*, *7*(5), e37027, 2012.

PÄÄKKÖNEN, P., PAKKALA, D. Reference architecture and classification of technologies, products and services for big data systems. Big Data Research, 2(4), 166-186, 2015.

PÁEZ, A., SCOTT, D. M., MORENCY, C. Measuring accessibility: positive and normative implementations of various accessibility indicators. Journal of Transport Geography, 25, 141-153, 2012.

ROLNIK, R., & KLINK, J. Crescimento econômico e desenvolvimento urbano: por que nossas cidades continuam tão precárias? *Novos estudos-CEBRAP*, (89), 89-109, 2011.

SCARINGELLA, R. S. A crise da mobilidade urbana em São Paulo. *São Paulo em perspectiva*, *15*(1), 55-59, 2001.

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=S0103-40142013000300004&lng=en&nrm=iso>. Acesso em 1 Junho de 2018.

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.

TRIBBY, C. P., ZANDBERGEN, P. A. High-resolution spatio-temporal modeling of public transit accessibility. Applied Geography, 34, 345-355, 2012.

WANG, M., & MU, L. Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA. *Computers, Environment and Urban Systems*, *67*, 169-175, 2018.

WILHEIM, J. Mobilidade urbana: um desafio paulistano. *Estudos avançados*, *27*(79), 7-26, 2013.

1. http://www.reuters.com/article/2010/11/05/ozatp-ghana-economy-idAFJOE6A40BG20101105 [↑](#footnote-ref-2)