

Análise de uma solução de rotas seguras para coletivos urbanos centrada no usuário

João Marcos A. M. Ramos¹, Thais R. M. B. Silva¹, Fabrício A. Silva¹

¹Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa (UFV)
Florestal - MG - Brazil

{joao.m.ramos, thais.braga, fabricio.asilva}@ufv.br

Resumo. *Estudos sobre fluxos de mobilidade vêm sendo cada vez mais relevantes devido ao aumento expressivo de dispositivos de coleta de dados e aplicações. Apesar disto, são poucos os trabalhos que abordam tal tema, além de ferramentas que auxiliam neste aspecto. Neste trabalho, foi implementada uma ferramenta que busca facilitar os estudos relacionados a fluxo de mobilidade, além de demonstrar sua aplicação em um estudo de caso, avaliando o impacto que a implementação de rotas seguras causa na cidade de São Paulo. Com os resultados, percebe-se que esta solução pode aperfeiçoar estudos já feitos na área de mobilidade, trazendo mais detalhe sobre as mudanças feitas, e como este impacto reflete nos deslocamentos.*

1. Introdução

A utilização intensiva de sensores capazes de rastrear a geolocalização em larga escala de usuários móveis trouxe consigo diversas possibilidades de estudos sobre o comportamento humano, visando a melhoria do dia-a-dia das pessoas. Uma das tarefas mais relevantes para o estudo da mobilidade humana corresponde ao entendimento de padrões de regularidade no movimento dos usuários, ou seja, a rotina dos mesmos [Barbosa et al. 2018]. Para empresas provedoras de serviços, o conhecimento sobre a regularidade de movimento de seus clientes traz vários benefícios, como a possibilidade de ofertar novos produtos, planejar a demanda por serviços em diferentes locais, planejar os melhores momentos para contactar os clientes, dentre outros.

Dentro da área de estudo de mobilidade, uma tarefa importante é a extração de fluxo de mobilidade, que consiste em analisar padrões de movimentos coletivos e determinar áreas ou locais que possuem uma alta movimentação. A análise do fluxo se mostra cada vez mais interessante para empresas e governos. Temos hoje diversas empresas que investem e contribuem cada vez mais com a área. Além disso, a percepção de fluxos para governos torna-se cada vez mais importante para o planejamento urbano, como a análise de movimentações de grandes metrópoles ou identificação de zonas de aglomerações de acordo com o horário do dia [Iqbal et al. 2014, Guo et al. 2012].

Um dos tipos de trabalho ligados à estudos de mobilidade são aqueles que investigam melhores opções de construção de rotas no cenário urbano, considerando fatores tais como clima, distância total a se percorrer ou ainda segurança. Tais trabalhos podem estar voltados à pedestres, ciclistas, veículos particulares ou ainda coletivos urbanos. É interessante notar como essas soluções podem se beneficiar de estudos de fluxos de mobilidade

em seu desenvolvimento, uma vez que podem considerar com melhor exatidão quais são, de fato, os deslocamentos de interesse da população.

Neste trabalho, será inicialmente apresentada uma solução para o cálculo de fluxo de mobilidades desenvolvida no contexto da biblioteca SENDAS. Em seguida, tal funcionalidade será utilizada na avaliação da construção de rotas seguras para ônibus urbanos proposta por [Almeida et al. 2021]. Esta estratégia de avaliação centrada no usuário, permitirá observar o impacto causado pelo uso de rotas seguras nos trechos em que a população da cidade de São Paulo realmente se desloca, em termos de distância percorrida e número de pontos de parada modificados.

O restante deste trabalho está organizado como segue: a seção 2 apresenta os trabalhos relacionados; a seção 3 descreve a implementação feita para a extração dos fluxos de mobilidade na ferramenta SENDAS. Um estudo de casos que demonstra o uso da funcionalidade de cálculo de fluxos implementada, bem como os resultados obtidos para o mesmo podem ser vistos nas seções 4 e 5. Por fim, na seção 6 são apresentadas as conclusões e alguns possíveis trabalhos futuros.

2. Trabalhos Relacionados

Estudos sobre o fluxo de mobilidade já vem sendo realizados por outros pesquisadores nos últimos anos. Neste contexto, temos trabalhos que tratam da importância e do uso do fluxo de mobilidade, além de ferramentas que auxiliam a identificação ou uso de fluxos.

[Guo et al. 2012] tratam sobre o agrupamento espacial de pontos GPS maciços para reconhecer lugares potencialmente significativos e a extração e mapeamento das medidas de fluxo dos agrupamentos para entender a distribuição espacial e tendências temporais dos movimentos. Este agrupamento é feito através da aglomeração espacial utilizando deslocamentos baseados em pontos de origem e destino. É apresentado um caso de estudos considerando o movimento de táxis em Shenzhen, China, onde foram observados os pontos de origem e destino de cada viagem de táxi por 5 dias. Os principais resultados obtidos pelos autores consistem numa variedade de padrões descobertos nos dados dos táxis, junto com a apresentação da abordagem para extrair informações de mobilidade origem-destino de grandes quantidades de dados. [Kon et al. 2021] descrevem a utilização do cálculo do fluxo de mobilidade para observar padrões de deslocamento em um sistema de compartilhamento de bicicletas, através de um método que pode ser usado para processar milhões de viagens. A proposta consiste em dividir a cidade em uma malha de várias regiões homogêneas utilizando uma grade uniforme e calcular o número de viagens de uma célula da grade para outra. Depois, são desenhados arcos direcionados para mostrar a direção dos fluxos e ajustar a origem e o destino destes de acordo com uma média ponderada baseada no uso da estação de coleta das bicicletas.

Dentre as ferramentas relacionadas, a biblioteca *MovingPandas*¹ implementa trajetórias, se mostrando assim mais interessante em situações onde os dados coletados são densos. No caso de dados esparsos, este tipo de representação pode apresentar várias falhas, tais como trajetórias incompletas ou deslocamentos improváveis. Outra opção de ferramenta é o *Scikit-Mobility*², que também permite a análise de trajetórias e de fluxos,

¹<https://anitagraser.github.io/movingpandas/>

²<https://scikit-mobility.github.io/scikit-mobility/>

corrige o problema da biblioteca citada acima, permitindo trabalhar com dados esparsos e completos. Porém, ambas as bibliotecas citadas não possuem mecanismos para otimizar o tratamento de grandes volumes de dados. Este trabalho se difere daqueles citados, uma vez que o módulo de fluxo do SENDAS foca em dados esparsos para deslocamentos de origem-destino, mas ainda assim, permite a análise com dados densos e análise de fluxos de trajetórias completas, permitindo que ocorra divisões e categorizações de acordo com o tempo, além de permitir a paralelização entre suas funções.

Trabalhos que abordam a construção de rotas seguras para ônibus urbanos também podem ser considerados relacionados à este, uma vez que este é o assunto abordado pelo estudo de caso desenvolvido. [Tompson et al. 2009] propõe *Hot Routes*, uma metodologia que mede a distribuição de risco do crime ao longo de uma rede linear, calculando a taxa de crimes por seção de estrada. Os autores demonstram a aplicação em dados criminais coletados das rotas de ônibus de Londres, Reino Unido. Este estudo mostra que a metodologia é fácil de implementar e versátil na sua aplicação, além de se mostrar mais eficaz que outras técnicas da literatura, visto que esta fornece uma visão mais localizada para análises de menor escala. [Almeida et al. 2021] busca avaliar uma solução de elaboração de rotas seguras em um contexto voltados para o transporte público rodoviário. Este trabalho será usado no estudo de casos e será explicado em maiores detalhes na seção 4. Ao contrário da abordagem do estudo de caso apresentado neste trabalho, a maneira como as análises são feitas nos trabalhos relacionados apresentados não leva em conta a análise do fluxo de deslocamento.

3. Sendas e Cálculo de Fluxo de Mobilidade

SENDAS(Scalable ENrichment for mobility DATA Sets) é uma biblioteca cujo objetivo consiste em facilitar os estudos de dados geoespaciais, inclusive, em relação à extração de fluxo de mobilidade. Os principais objetivos e vantagens desta biblioteca consistem em ser eficiente no que diz respeito ao uso de memória, possuir um bom desempenho computacional em termos de processamento, além de permitir operações paralelizáveis e personalizáveis. O SENDAS é um projeto do NESPED/UFV-Campus Florestal³, que atualmente se encontra em estado de desenvolvimento e, em breve, estará disponível para a comunidade acadêmica que trabalha com análise de dados geoespaciais.

Para o desenvolvimento da biblioteca foram selecionado uma linguagem de programação e um *framework* que se adéquem aos objetivos do projeto, priorizando o desempenho e escalabilidade. Para isto, foi escolhido o *framework* Spark⁴, que, segundo [Damji et al. 2020], possui como principais filosofias velocidade, facilidade de uso, modularidade e extensibilidade, além de permitir a execução em *clusters*. Em relação à linguagem, optou-se pelo uso da linguagem Scala, devido aos ganhos de desempenho e expressividade, além de se apresentar uma linguagem com um crescimento significativo de utilização na área de ciência de dados.

Para entendermos melhor a utilização da biblioteca SENDAS, podemos observar a figura 1, onde dividimos o funcionamento da biblioteca em etapas, sendo estas as etapas de entrada dos dados, pré-processamento dos mesmos e utilização das funcionalidades. Na etapa de entrada, a biblioteca recebe um conjunto de pontos geolocalizados repre-

³<https://nesped.caf.ufv.br/>

⁴<https://spark.apache.org/>

sentados pela latitude e longitude sendo, que esta consiste na representação de espaço, e alguma coluna de unidade de tempo, podendo ser por exemplo uma data ou hora. Estes dados devem ser inseridos através de um arquivo de entrada, podendo ser tanto qualquer formato reconhecido pelo Spark, como um CSV, Parquet, XML ou Json.

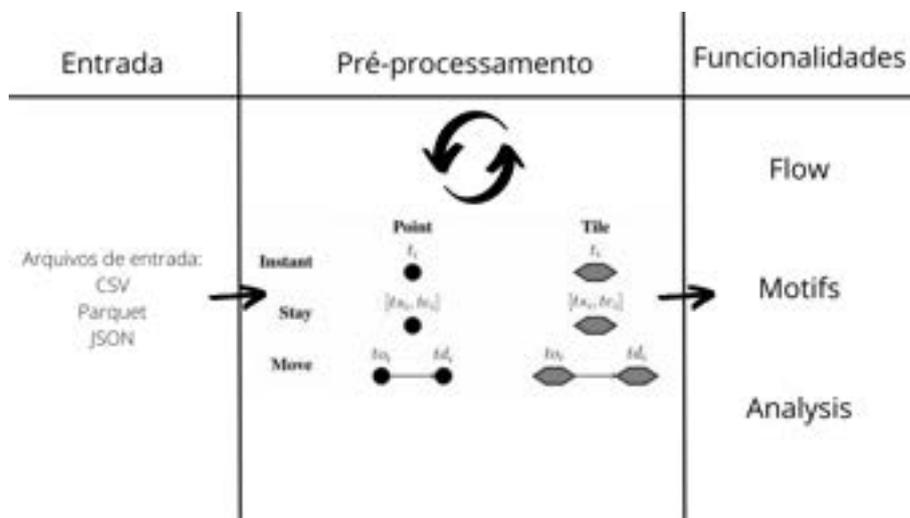


Figura 1. Tabela das etapas de utilizações do SENDAS

Já na etapa de pré-processamento, o objetivo é preparar os dados da entrada e transformá-los, para a utilização dos algoritmos ou funcionalidades fornecidos pela biblioteca. Nesta etapa, podemos escolher uma das visões de trajetória disponibilizadas. Existem três categorias: *Instant*, representa uma visão espacial observada durante um certo instante de tempo, *Stay* representa uma estadia ao longo de um determinado período de tempo e o *Move* representa um deslocamento entre duas localizações. Para representação espacial dos dados, pode-se escolher entre ponto (*point*) ou região (*tile*). A partir disto, podemos obter diversas saídas que podem ser utilizadas em outras partes do código ou na etapa de funcionalidades, bem como ainda externalizá-las como um arquivo dos mesmo tipos descritos para as entradas.

Na última etapa, temos as funcionalidades, que consistem nos principais algoritmos ofertados pela biblioteca. Entre estas funcionalidades, temos o cálculo do fluxo, em que é necessário passarmos por quatro camadas: *StayPoints*, Tesselações, Deslocamentos e o Fluxo de Deslocamentos. Os três primeiros estão brevemente descritos a seguir. O último é o principal objetivo de estudo deste trabalho e será descrito em maiores detalhes na seção 3.1.



Figura 2. Exemplo dos dados em cada camada

- **Camada de *StayPoints*** - Essa camada conta com ferramentas para identificar todos os deslocamentos individuais que ocorreram através do algoritmo de identificação de *StayPoints* descrito em [Montoliu et al. 2011]. Este algoritmo utiliza dos dados de entrada, detectando locais (*StayPoints*) para as quais foi observado um tempo de permanência maior que um limite informado pelo usuário. *StayPoints* podem ser úteis para encontrar lugares de interesse ou encontrar padrões de deslocamentos. A utilização do algoritmo de *StayPoints* se torna vantajosa, pois filtra a quantidade de pontos intermediários, caso os dados sejam densos, que não representam os fins e começos dos deslocamentos. Podemos observar que a figura 2 (a) ilustra a identificação de *StayPoints* a partir dos pontos geolocalizados passados como entrada. Os *StayPoints* estão marcados em vermelho na imagem.
- **Camada das Tesselações** - Uma tesselação pode ser definida como um particionamento do espaço, sendo que cada partição corresponde a uma região. O foco é atribuir um identificador a cada partição com o objetivo de referenciá-lo através do mesmo e não mais da região em si. A tesselação define uma abstração de como fazer a conversão entre pontos e regiões, e pode ser implementada de formas diversas. Por exemplo, o H3⁵, que consiste num sistema de indexação geoespacial que utiliza de hexágonos para definir as regiões de interesse, sendo que o tamanho de cada hexágono varia de acordo com a resolução utilizada na hora da conversão. Também, é possível que o próprio usuário fornecer uma base de dados com os polígonos com os identificadores associados, reforçando a ideia do *tesselation* de ser extensiva. Podemos observar que a figura 2 (b) ilustra a aplicação das tesselações para identificarmos as regiões a partir dos *StayPoints* passados como entrada. Cada região está marcada de verde na imagem.
- **Camada dos Deslocamentos** - Esta camada consiste na identificação dos deslocamentos individuais de cada trajetória. Assume-se que ocorre um deslocamento quando ocorrem dois registros do mesmo identificador em duas regiões diferentes. Podemos observar que a figura 2 (c) ilustra a identificação dos deslocamentos para identificarmos as regiões a partir dos pontos *StayPoints* passados como entrada. Cada linha representa um deslocamento entre uma região de origem até uma região de destino, sendo que a origem consiste na cor vermelha e o destino corresponde a cor azul na imagem.

3.1. Camada de Fluxos de Deslocamentos

Definimos um fluxo como a agregação de deslocamentos que aconteceram entre duas regiões A e B em um determinado intervalo de tempo. Caso não se apresente uma divisão de tempo, consiste nos deslocamentos totais entre as regiões. O objetivo desta camada é o cálculo dos fluxos, que baseiam-se nos deslocamentos individuais apresentados, agregando-os, e assim, indicando o fluxo de deslocamento entre regiões.

A implementação desta camada espera como entrada uma representação espacial de região, junto com uma visão de trajetória de movimento. Com base nestas entradas, os

⁵<https://eng.uber.com/h3/>

dados são agregados baseando-se nos identificadores de origem e de destino que demonstram um deslocamento, e nesta agregação, são identificados e contabilizados os fluxos, baseados nas origens e destinos.

Pode-se também configurar tanto uma divisão de tempo quanto uma categorização de tempo para o cálculo do fluxo. No caso da divisão do tempo, utiliza-se para dividir o tempo total em unidades de tempo, sendo muito útil caso seja necessário analisar um agregado do conjunto inteiro, como por exemplo, um fluxo por dia ou por mês. Já na categorização, busca-se para filtrar o tempo, como pode ser observado na Figura 3. A diferença entre as duas, consiste que, enquanto a divisão de tempo apenas particiona o tempo total da base de dados em lacunas ou intervalos de tempo, as categorias permitem uma filtragem semântica do tempo, de acordo com faixas estabelecidas pelo usuário.

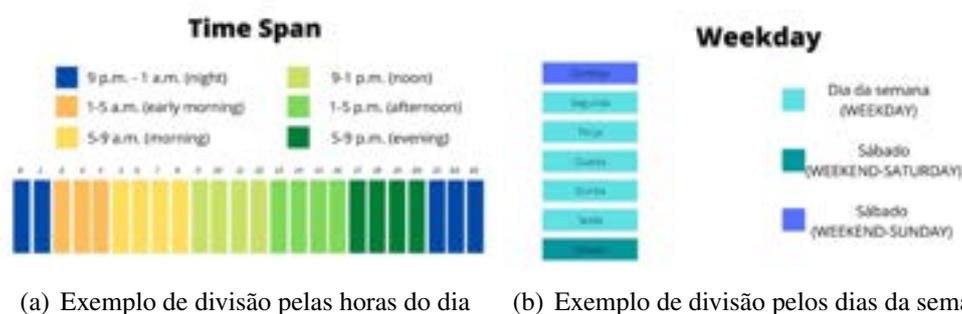


Figura 3. Categorização de Fluxo de Deslocamentos

Uma vez que SENDAS pretende ser uma biblioteca extensível, ela também oferece uma possibilidade do usuário criar seu próprio método de caracterização, bastando informar em qual unidade de tempo será avaliada, podendo ser tanto dias, horas, semanas, e como serão divididos os intervalos destas unidades de tempo. Podemos ver nas Figuras 3(a) e 3(b) exemplos de categorizações feitas, por hora e por dia da semana respectivamente, as quais já são disponibilizadas pela biblioteca. Estas divisões podem ser utilizadas para encontrar certos padrões de deslocamentos baseados em determinados horários ou dias, além de permitir melhor organização durante o processo de análise e filtragem de determinados períodos específicos sem a necessidade de operações muito complexas.

Por fim, teremos a saída, que consiste em duas possibilidades: Uma matriz origem-destino e uma matriz de regiões. A matriz de origem-destino nos permite uma análise de cada fluxo, e neste contexto, o *inflow* representa a quantidade de fluxos que estão chegando na região de destino, enquanto o *outflow* representa a quantidade de fluxo que está saindo da região de origem. Já a representação por região, permite a visualização de acordo com cada região passada, e os dois tipos de fluxos anteriormente se referem aos deslocamentos que chegam e que saem da mesma região. Em ambos os casos tanto *inflow* quanto *outflow* se aplicam em cada intervalo de tempo, caso seja usado.

É possível utilizar os dados tanto da matriz origem-destino, quanto da matriz de regiões para calcularmos as seguintes métricas que podem auxiliar na análise de dados: fluxo total, fluxo médio e fluxo líquido.

O fluxo total, dado pela equação $F_{total} = F_{in} + F_{out}$ indica a soma dos fluxos de inflow (F_{in}) e outflow (F_{out}) em uma determinada região.

O fluxo líquido, que é dado pela equação $F_{net} = F_{in} - F_{out}$ indica a quantidade de fluxos que passam em uma determinada região. Se este valor está positivo, indica que um determinado local recebe mais fluxo do que sai; caso este número esteja negativo, indica que o fluxo de saída deste lugar é maior que o de entrada.

Já o fluxo médio indica a quantidade média de fluxo da região, podendo ser observada durante o intervalo de tempo definido. Sua fórmula consiste em $F_{avg} = \frac{F}{t}$, onde t é a quantidade de intervalos de tempo definidos e F podendo ser qualquer tipo de fluxo.

4. Estudo de Caso: Rotas Seguras para Ônibus e Fluxos de Mobilidade

Como forma de exemplificar o uso da funcionalidade de cálculo de fluxos de mobilidade oferecida por Sendas, neste trabalho apresentaremos um estudo de caso envolvendo sua aplicação. Em particular, o cálculo de fluxos será utilizado como parte de uma estratégia de avaliação de uma solução que pretende oferecer rotas seguras para veículos coletivos urbanos na cidade de São Paulo.

Uma cidade, em geral, possui diversas linhas de ônibus distintas, sendo que cada uma delas pode percorrer uma ou mais rotas entre sua origem e destino. Em particular, para cada linha, existem as opções de rota mais segura e rota mais curta. Cada rota é composta por uma sequência de itens, que correspondem, neste trabalho, a cada esquina por onde o ônibus deve passar. Além disto, determinados itens são pontos de paradas, que consistem em locais onde ocorre o embarque e desembarque de passageiros.

Tradicionalmente, um ônibus tende a seguir por uma rota mais curta, isto é, aquela que, considerando as limitações das vias da cidade, percorrer a menor distância entre cada um dos pontos de parada que compõem a rota. Porém, não necessariamente a rota mais curta é segura, isto é, evita áreas que possuem altos índices de criminalidade. Muitas vezes, neste contexto, o cidadão fica exposto à violência tanto dentro dos coletivos, como nos pontos de parada. Desta forma, uma alternativa seria o uso de rotas seguras, as quais evitam áreas criminais tanto quanto possível.

Dentre diversos trabalhos propostos na literatura para o cálculo de rotas seguras visando veículos particulares, a solução de [Almeida et al. 2021] é uma das poucas alternativas voltadas para ônibus urbanos. Nela, foi criada uma solução que calcula a rota mais segura para cada linha de ônibus da cidade de São Paulo. Além de sugerir caminhos que evitam locais com altos índices de criminalidade, a solução ainda visa alterar determinados pontos de paradas presentes em zonas com alta periculosidade para outros mais próximos e menos perigosos. A solução de rotas seguras abordadas por este trabalho foi avaliada considerando integralmente cada linha de ônibus e verificando o aumento da segurança e do deslocamento quanto comparado à rota mais curta. Porém, esse trabalho não considerou a maneira como a população se desloca pela cidade e não avaliou o impacto para esta na mudança dos trajetos e dos pontos de parada.

Em relação a solução no trabalho de [Almeida et al. 2021], funciona da seguinte maneira: inicialmente, identifica os *clusters* criminais com o algoritmo *DensityBased Spatial Clustering of Applications with Noise (DBSCAN)*. É interessante frisar que para uma zona ser considerada de alta criminalidade, só são contabilizados crimes de furto, homicídio simples, roubo e tráfico de entorpecentes, e apresentar pelo menos 3 crimes por mês por média em um raio de até 100 metros. Depois de obter os *clusters* criminais, usa-se

a ferramenta OSMnx⁶ para construir um grafo representando a cidade de São Paulo, onde cada esquina da cidade é representada como um vértice, e as ligações entre elas, arestas direcionais. Além do grafo, são feitos os polígonos baseados em cada zona criminal, para delimitá-las. A partir do centro do polígono é verificado se cada vértice está dentro ou não do mesmo; caso esteja, o vértice é considerado perigoso. Por último, é utilizado um outro conjunto de dados com informações das rotas de ônibus da cidade de São Paulo, e são colocados os pontos de parada no grafo. Esta colocação é feita procurando os itens mais próximos utilizando a distância euclidiana, e classificando-os em pontos de parada.

O próximo processo no trabalho é o de identificação das rotas mais curta e mais segura para cada linha de ônibus. A própria ferramenta OSMnx faz a identificação das rotas mais curtas. Porém, para as rotas seguras, o processo é feito entre pares de vértices, sendo que em cada par são geradas todas as rotas possíveis de um vértice ao outro, e para cada elemento desta rota utiliza-se uma fórmula que calcula o quão próximo ele se encontra de um *cluster* criminal. Caso ele esteja muito próximo, este vértice é penalizado na escolha da rota. Desta forma, divide-se a rota em trechos menores para que haja um detalhamento maior para cada par de pontos de parada consecutivos.

Ampliando ainda mais a solução segura, [Almeida et al. 2021] fornece uma solução de remover determinados pontos de paradas para aumentar a segurança da rota. Um ponto de parada só é removido se existe um outro ponto fora de uma região criminal, e localizado em até um raio de 250m, além de que siga uma restrição máxima de distância real de 1000 metros. Depois de aplicar todas as etapas acima, os dois tipos de rotas (curta e segura) são gerados para cada linha de ônibus, sendo comparadas em termos de distância total percorrida e nível de segurança. Entretanto, vale observar que tal avaliação é feita sempre considerando integralmente cada linha de ônibus, isto é, desde seu ponto de parada inicial até o final, e sem considerar o impacto real sobre a população, visto não serem utilizados os fluxos de deslocamento desta.

A alteração de paradas pode afetar o fluxo de mobilidade da população, pois um determinado ponto desse tipo, apesar de ser perigoso, pode ser importante. Da mesma forma, pode-se alterar alguns pontos de parada aumentando significativamente a distância percorrida por alguns passageiros, aumentando o tempo de viagem e prejudicando aqueles que passam por tal rota. Outro fator relevante é que as linhas de ônibus não necessariamente refletem o fluxo da mobilidade. Uma determinada linha pode ter trechos com maior embarque e desembarque de passageiros. Sendo assim, determinados trechos da rota são mais relevantes para a população do que outros.

Para encontrarmos quais trechos são mais relevantes, utilizamos a função de identificação de fluxos da biblioteca Sendas, para que assim, identifiquemos os fluxos de deslocamentos reais da população. Encontrando de quais pontos de parada as pessoas estão partindo e quais são seus destinos, é possível identificar quais trechos das linhas de ônibus são os mais relevantes e, caso sejam alterados, o quão impactante foi tal alteração. O objetivo, ao final, é comparar o uso, pela população, dos trechos identificados como relevantes por meio da rota mais curta e da rota mais segura. Essa estratégia de avaliação conseguirá observar melhor o impacto de uma mudança de rota pelos coletivos, do que aquela que observa todas as linhas de ônibus consideradas por inteiro.

⁶<https://osmnx.readthedocs.io/en/stable/>

Para chegarmos aos resultados pretendidos, foi passada como entrada para a biblioteca SENDAS uma base de dados contendo pontos geolocalizados da cidade de São Paulo. Essa base possui 11.351.545 registros que contém deslocamentos em 266 dias distintos no período de 2021. Foram observados 356.725 usuários, sendo que grande parte dos mesmos apresentam em sua maioria, menos de 200 aparições nos dados.

H3_hash	App_user_id	Date_time	Cod_município	Latitude	Longitude
baa8100088b7ff	90000000	2021-12-05T11:56:38.000-03:00	3550308	-23.67	-46.58
baa810011737ff	90000000	2021-12-05T16:26:21.000-03:00	3513801	-23.69	-46.59
baa81001c227ff	90000000	2021-12-04T13:28:25.000-03:00	3513801	-23.67	-46.62
baa8100625b7ff	90000000	2021-12-05T16:16:08.000-03:00	3550308	-23.63	-46.45
baa8100912d8ff	90000000	2021-12-05T14:35:11.000-03:00	3550308	-23.61	-46.44
baa8100b3098ff	90000000	2021-12-04T18:53:01.000-03:00	3529401	-23.66	-46.44
baa8100cb987ff	90000000	2021-12-04T20:08:13.000-03:00	3550308	-23.5	-46.67
baa810316547ff	90000000	2021-12-04T21:02:53.000-03:00	3510609	-23.55	-46.54
baa810770557ff	90000000	2021-12-05T14:15:13.000-03:00	3550308	-23.45	-46.56
baa8100a12967ff	90000000	2021-12-05T13:47:26.000-03:00	3550308	-23.53	-46.54
baa810395347ff	90000000	2021-12-05T09:44:21.000-03:00	3550308	-23.47	-46.74
baa810306357ff	90000000	2021-12-05T13:38:30.000-03:00	3522505	-23.55	-46.92
baa8103108d8ff	90000000	2021-12-04T18:56:10.000-03:00	3505708	-23.54	-46.87
baa810702997ff	90000000	2021-12-05T15:28:52.000-03:00	3550308	-23.49	-46.45
baa81004d1a8ff	90000000	2021-12-04T14:45:20.000-03:00	3513009	-23.59	-46.84
baa8100a06a8ff	90000000	2021-12-05T05:02:45.000-03:00	3547809	-23.7	-46.51
baa8100af247ff	90000000	2021-12-05T13:00:14.000-03:00	3547809	-23.7	-46.49
baa810214327ff	90000000	2021-12-05T14:01:24.000-03:00	3509305	-23.41	-46.67
baa810396168ff	90000000	2021-12-05T12:59:36.000-03:00	3550308	-23.48	-46.7
baa810006b17ff	90000000	2021-12-05T08:22:25.000-03:00	3513801	-23.73	-46.62

Figura 4. Amostra dos dados

Ao primeiro momento, vamos aplicar o algoritmo de detecção de *StayPoints*, descritos nas camadas acima, visto que certas aparições nos dados podem simular deslocamentos curtos que provavelmente não iriam necessitar da utilização de ônibus, poluindo assim, os fluxos.

Após a aplicação, precisa-se converter os dados da visão de pontos para a de região, visto que a *feature* de fluxo utiliza desta forma. Para isto, utilizamos a camada de tesselações para convertermos para setores censitários, que consistem em uma unidade territorial demarcada pelo IBGE para fins de organizar a coleta de dados de suas pesquisas domiciliares, visto que, estes setores apresentam uma representação melhor da cidade, dos bairros e das subdivisões físicas, mais acuradas que o próprio H3 em relação a cidade de São Paulo. Além disso, para facilitar a visualização, foi adicionado depois os formatos das geometrias dos setores. É necessário também converter a visão para deslocamentos assim como foi descrito nas camadas. Visto que, com todos os deslocamentos individuais calculados, agora pode-se calcular o fluxo. O fluxo foi calculado sem nenhuma categorização ou divisão de tempo, visto que não há necessidade de dividir os deslocamentos por unidades de tempo, já que o interesse deste estudo de caso consiste em descobrir todos os fluxos considerando todo o intervalo de tempo dos dados de entrada.

Com isso, foram obtidos 58.581 deslocamento entre as regiões. Porém, há alguns deslocamentos que são dentro do próprio setor, visto que são deslocamentos curtos, que normalmente não haveria necessidade de utilização do ônibus para realizá-los, por-

tanto, sendo removidos. Com isto, obtivemos 56.776 fluxos. Com a ferramenta Kepler⁷, podemos observar melhor através do mapa descrito na Figura 5. Sendo que a cor azul representa a origem e vermelho o destino.



Figura 5. Visualização através da ferramenta Kepler de uma parte dos fluxos de São Paulo.

Foi possível observar que grande parte dos fluxos são individuais. Temos também que a média de pessoas nos fluxos é 3.68, seu maior valor é 292, e seu menor valor é 1, que também é o valor mais frequente.

Cada fluxo de deslocamento identificado possui um setor censitário de origem e outro de destino. O próximo passo consiste em buscar, para cada fluxo, um trecho de linha de ônibus que o interligue. Para a escolha, busca-se aquele que possui o menor número de pontos de parada entre os setores censitários de origem e destino. Para obter este resultado, é necessário identificar quais são os possíveis ônibus que realizam cada deslocamento baseado no setor de origem e de destino. Para isto, foi considerada a mesma base de rotas de ônibus utilizadas por [Almeida et al. 2021], para que assim, seja possível obter os mesmos resultados. Foi necessário identificar em qual setor censitário cada ponto de origem e destino se localizam, utilizando a camada de tesselações novamente neste processo. Sabendo a quais setores cada ponto pertence, pode-se descobrir quais são os trechos de linhas possíveis para cada deslocamento. Baseado nos setores de origem e destino, são obtidos quais trechos de linha cumprem aquele caminho. Busca-se as linhas que atendem da origem até destino, e as ordena baseado no número de pontos de paradas entre os dois escolhendo a primeira que atende esta característica. A partir destas etapas, é possível observar cada trecho de linha de ônibus que atende cada fluxo.

Para cada trecho de linha escolhido para os fluxos identificados, foram geradas, conforme o trabalho de [Almeida et al. 2021], a rota mais curta e rota mais segura. Esses dois tipos de rota foram comparados em termos de distância total percorrida, número de vértices alterados e quantidade de pessoas impactadas.

Para alguns trechos, ao se calcular a rota mais segura de acordo com [Almeida et al. 2021], pontos de parada de destino e/ou origem foram removidos. Estes casos são analisados de maneira diferente. Antes da etapa de identificação de rota,

⁷<https://kepler.gl/>

os pontos de origem e destino são convertidos para, caso alterados, seus novos valores, e caso de remoção, para outro ponto de parada.

5. Resultados

No desenvolvimento deste trabalho, foi utilizado somente um notebook Acer Aspire 5 a514-54 com o sistema operacional Windows 10, processador i7-8550U, com 12GB de RAM e 240 GB de SSD. Ele foi capaz de executar todas as etapas necessárias.

Para esta análise, consideramos algumas métricas. A principal a ser analisada é a quantidade de pessoas em cada rota. Temos também, a distância percorrida, as métricas relacionadas aos vértices e pontos de parada. Todas estas métricas serão descritas mais a fundo em suas respectivas sessões abaixo.

5.1. Distância Percorrida e Número de Vértices

A Distância Percorrida consiste no tamanho do trajeto entre os pontos de origem e destino dos fluxos, não da linha como um todo. Portanto, é a distância viajada por cada pessoa no fluxo. Temos também, a distância extra percorrida, que é dada pela subtração do tamanho da rota segura, que é sempre maior ou igual, pelo tamanho da rota mais curta. Este valor nunca se torna negativo, e o 0 significa que a rota segura não sofreu alteração de tamanho em relação a rota mais curta. Já o número de vértices consiste no número de pontos de parada e esquinas que compõe a rota daquele específico trecho da linha de ônibus.

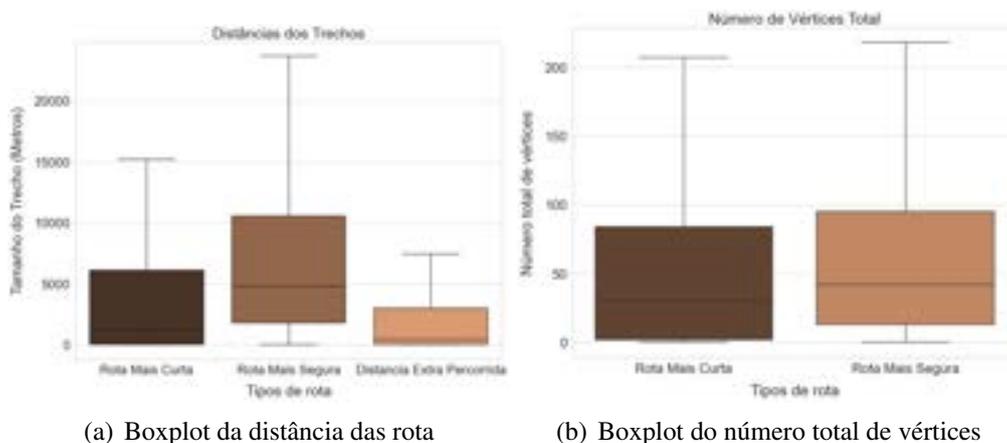


Figura 6. Boxplots do Número Total de Vértices e Tamanho da rota com os outliers ocultos

Podemos observar a distância das rotas na Figura 6 (a). Observamos que, as rotas seguras são significativamente maiores que as rotas mais curtas, sendo que a mediana das rotas curtas está bem abaixo dos 5km, enquanto nas rotas seguras, estão bem próximas dos 5km. Até mesmo no tamanho máximo, podemos observar este aumento significativo. Porém, se observarmos a distância extra percorrida, vemos que grande parte dos aumentos ocorre na faixa de 1km. Em relação ao número total de vértices, vemos sua aplicação na Figura 6 (b), e podemos observar que o aumento de vértices não chega a ser tão significativo quanto a distância, apesar da haver um aumento.

Medidas	N. Vértices Diferentes	N. Pontos de Parada Diferente	N. Vértices Iguais	N. Pontos de Parada Iguais
Média	15.491141	3.760339	45.751427	6.390799
Mínimo	0.000000	0.000000	0.000000	0.000000
Máximo	547.000000	83.000000	493.000000	90.000000
25%	0.000000	0.000000	1.000000	0.000000
50%	4.000000	0.000000	24.000000	1.000000
75%	13.000000	2.000000	66.000000	9.000000
Desvio Padrão	35.720046	8.654518	59.515003	10.458669

Tabela 1. Tabela dos vértices e pontos de parada iguais e diferentes.

5.2. Vértices e Pontos de Parada Iguais e Diferentes

Os vértices e pontos de parada foram descritos no seção 4. Porém, após a realocação de determinados pontos de parada, vértices podem ser adicionados e retirados. Além disso, determinados pontos de parada podem ser alterados do trecho original. Para medirmos o impacto de tais alterações, comparamos os números de pontos de paradas e vértices das rotas mais curtas e das rotas mais seguras, como podemos ver na Tabela 1.

Podemos observar que, em relação a vértices, temos que em média, alteram-se poucos vértices do trecho original. Podemos observar também que, em determinados trechos de rota, não há alteração nenhuma, visto que a rota mais curta já é uma rota segura, além de que, grande parte das alterações estão em até 4 vértices diferentes do trecho original. Já em relação aos pontos de paradas, vemos que grande parte dos trechos não alteram pontos de parada, ou seja, mantém a integridade do trecho original.

5.3. Fluxo Versus Distância, Vértices e Pontos de Parada Realocados

Para entendermos o número de pessoas impactadas pelas rotas, podemos olhar pelo fluxo, sendo que cada fluxo corresponde ao agregamento de deslocamentos individuais. Para isto basta considerarmos que cada deslocamento corresponde a uma pessoa, sendo assim, teremos o número de pessoas.

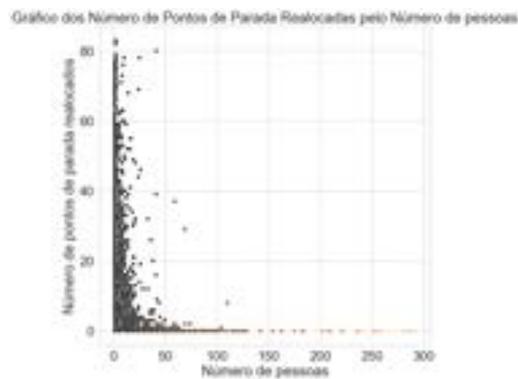
Podemos ver na Figura 7 (a) o resultado por analisar o número de pessoas no fluxo, e a distância adicional que cada linha de ônibus irá percorrer, devido as realocações, junto com o número de pessoas afetadas por esta mudança. Já na Figura 7 (b), temos o resultado da análise em relação ao número de vértices adicionados na rota pelo número de pessoas. Por último, na Figura 7 (c) temos a observação do número de pessoas que foram afetadas pelas realocações, e o número de pontos parada que foram realocados no trecho da rota. Podemos observar que grande parte das mudanças significativas impactam menos de 50 pessoas, enquanto os de maiores impactos afetam uma parcela bem menor das pessoas.

6. Conclusão e Trabalhos Futuros

Este trabalho apresentou conceitos relacionados ao fluxo de mobilidade, assim como, a importância do seu cálculo. Também foi apresentada uma implementação feita na ferramenta SENDAS para facilitar estudos relacionados a esta área. Por último, foi realizada uma análise conjunta dos padrões de mobilidade da cidade de São Paulo e das novas trajetórias propostas por uma solução de construção de rotas seguras [Almeida et al. 2021],



(a) Gráfico da distância percorrida pelo número (b) Gráfico do número de vértices no trecho de pessoas *StayPoints* pelo número de pessoas



(c) Gráfico do número de vértices realocados pelo número de pessoas

Figura 7. Gráficos em relação ao número de pessoas

buscando trazer informações a cerca do impacto da primeira sobre a segunda. Conforme foi apresentado nos resultados do estudo de casos, podemos perceber que grande parte dos impactos causados pelas alterações são impactos pouco significativos, afetando pequenas parcelas da população ou mudanças em pequenas distâncias. Sobre trabalhos futuros, busca-se incrementar a ferramenta implementada com novas funcionalidades em relação a fluxo de mobilidade, ou facilitar o cálculo com novas métricas. Também é esperada a realização de outros casos de estudo levando em consideração outras cidades.

Referências

- Almeida, V. G. J., Silva, T. R. M. B., and Silva, F. A. (2021). Avaliação do uso de um modelo de criminalidade na determinação de rotas seguras para Ônibus urbanos. Monografia (Bacharel em Ciência da Computação), UFV (Universidade Federal de Viçosa), Florestal, Minas Gerais, Brasil.
- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., and Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734:1–74. Human mobility: Models and applications.

- Damji, J., Wenig, B., Lee, D., Das, T., and Safari, a. O. M. C. (2020). *Learning Spark, 2nd Edition*. O'Reilly Media, Incorporated.
- Guo, D., Zhu, X., Jin, H., Gao, P., and Andris, C. (2012). Discovering spatial patterns in origin-destination mobility data. *Transactions in GIS*, 16.
- Iqbal, M. S., Choudhury, C. F., Wang, P., and González, M. C. (2014). Development of origin-destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74.
- Kon, F., Ferreira, , Souza, H., Duarte, F., Santi, P., and Ratti, C. (2021). Abstracting mobility flows from bike-sharing systems. *Public Transport*.
- Montoliu, R., Blom, J., and Gática-Pérez, D. (2011). Discovering places of interest in everyday life from smartphone data. *Multimedia Tools and Applications*, 62:179–207.
- Tompson, L., Partridge, H., and Shepherd, N. (2009). Hot routes: Developing a new technique for the spatial analysis of crime. *Crime Mapping: A Journal of Research and Practice*, 1:2009.