

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

João Marcos A. M. Ramos¹, Vinícius G. J. Almeida¹, Henrique S. Santana¹,
Thais R. M. B. Silva¹, Fabrício A. Silva¹

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

{joao.m.ramos, vinicius.jesus, henrique.s.santana}@ufv.br

{thais.braga, fabricio.asilva}@ufv.br

Resumo. *A mobilidade ciente de contexto tem o potencial de tornar a maneira como nos deslocamos mais eficiente, segura e sustentável. Dentre os contextos possíveis, a segurança, em termos de níveis de criminalidade das regiões, é um que tem sido utilizado visando o cálculo de rotas mais seguras. Porém, os trabalhos existentes focam em veículos particulares, e não avaliam o impacto gerado para os cidadãos como um todo. Neste trabalho, uma solução existente para o cálculo de rotas seguras é avaliada no contexto de transporte coletivo de ônibus em termos do impacto causado para os passageiros. Os resultados mostraram que, em geral, a mudança de uma rota de ônibus para se tornar mais segura aumenta o trecho viajado em poucos quilômetros para a maioria dos passageiros. Além desta análise, uma ferramenta para extração de fluxos de mobilidade escalável foi também implementada como contribuição do trabalho.*

Abstract. *Context-aware mobility has the potential to make the way we travel more efficient, safer and more sustainable. Among the possible contexts, security, in terms of crime levels in the regions, is one that has been used to calculate safer routes. However, the existing works focus on private vehicles, and do not assess the impact generated for citizens as a whole. In this work, a existing solution for calculating safe routes is evaluated in the context of public bus transport in terms of the impact caused to passengers. The results showed that, in general, changing a bus route to make it safer increases the distance traveled by a few kilometers for most passengers. In addition to this analysis, a tool for extracting scalable mobility flows was also implemented.*

1. Introdução

A mobilidade ciente de contexto é uma área que vem crescendo nos últimos anos, e visa utilizar dados sobre usuários e o ambiente para melhorar a maneira como as pessoas se deslocam [Santos et al. 2017]. Com isso, espera-se que o deslocamento dos cidadãos seja mais eficiente, seguro, sustentável e personalizado. Dentre os contextos que podem ser utilizados, a segurança é um que tem ganhado destaque. Neste caso, o objetivo é traçar rotas de deslocamento que sejam mais seguras, em termos dos níveis de criminalidade das regiões por onde o veículo ou as próprias pessoas irão passar, para proteger os usuários.

Os trabalhos existentes até o momento têm direcionado os esforços para soluções voltadas a veículos particulares. Em geral, o objetivo desses trabalhos é traçar uma rota segura para um carro ou moto que deseja se deslocar de um local de origem para um de destino. Essas propostas, apesar de serem relevantes para a mobilidade, não foram avaliadas no contexto de transporte coletivo de ônibus. Nas grandes cidades, os veículos coletivos são responsáveis por transportar boa parte dos cidadãos. Portanto, a ciência de contexto de criminalidade também deve ser incorporada nas rotas dos ônibus para que reflita na segurança destes passageiros.

O único trabalho encontrado que visa a construção de rotas seguras para o transporte coletivo, em particular ônibus, é o [Vinícius G. J. Almeida 2022]. Essa solução contempla o mapeamento de níveis de criminalidade em regiões com formatos diversos, apresentando uma fórmula objetiva para o cálculo desse nível. Essa solução foi implementada para as linhas de ônibus da cidade de São Paulo/SP com o intuito de alterar os caminhos por onde os ônibus passam entre uma parada e outra, para que sejam mais seguros. Porém, esse estudo apenas avaliou o impacto em termos das rotas, sem olhar para a perspectiva dos passageiros.

O objetivo deste trabalho é fazer uma avaliação centrada no usuário da solução de [Vinícius G. J. Almeida 2022]. Foi avaliado o impacto das alterações nas rotas para os usuários, tendo como base um conjunto de dados reais de mais de 300.000 usuários móveis. Os resultados mostram que as alterações nos caminhos afetam, em geral, um número pequeno de cidadãos, que precisam se deslocar por até 1 km a mais no ônibus para chegarem ao seus destinos.

Para que os resultados pretendidos fossem coletados, foi preciso extrair os fluxos de deslocamentos desses milhares de usuários considerados, mapeando, em seguida, esses fluxos em trechos de linhas de ônibus. Para isso, foi necessário desenvolver uma ferramenta para extração de fluxo que seja escalável. Essa ferramenta conseguiu melhorar em até 50% o tempo de processamento, se comparada com uma solução disponível na literatura. Esta é a segunda contribuição deste trabalho.

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. A avaliação da solução de rotas seguras no contexto de ônibus é descrita na Seção 4. Por fim, na Seção 5 são apresentadas as conclusões e alguns possíveis trabalhos futuros.

2. Trabalhos Relacionados

Os trabalhos relacionados ao presente estudo podem ser organizados em duas partes, de acordo com as suas duas contribuições: análise de rotas seguras e extração de fluxos.

Em relação à análise de rotas seguras, existem alguns trabalhos na literatura, tais como [Ladeira et al. 2020, de Souza and Villas 2020, Ladeira et al. 2019]. Porém, todos eles possuem foco em veículos particulares, e não analisam o impacto coletivo na mobilidade. O único trabalho encontrado sobre esse assunto para ônibus é o [Vinícius G. J. Almeida 2022]. Esse trabalho estende o [Ladeira et al. 2019], ajustando as funcionalidades para o transporte coletivo de ônibus. Inicialmente, os autores apenas avaliaram o impacto das alterações dos caminhos nas características das rotas, sem olhar

no impacto causado nos passageiros. Essa solução foi escolhida para ser avaliada neste trabalho atual.

Em relação aos fluxos de mobilidade, os autores de [Guo et al. 2012] tratam sobre o agrupamento espacial de pontos GPS massivos 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. Já os autores de [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.

Também existem duas ferramentas conhecidas para a extração de fluxos. A biblioteca *MovingPandas*¹ trata de trajetórias, se mostrando mais interessante em situações onde os dados coletados são densos. No caso de dados esparsos, este tipo de representação pode apresentar falhas, tais como trajetórias incompletas ou deslocamentos improváveis. Outra opção de ferramenta é o Scikit-Mobility², que também permite a análise de trajetórias e de fluxos, corrigindo o problema da biblioteca citada acima, uma vez que permite trabalhar com dados esparsos. Porém, ambas as bibliotecas citadas não possuem bom desempenho para grandes volumes de dados, como a proposta utiliza neste trabalho.

3. Identificação do Fluxo de Mobilidade

Dentro da área de estudo de mobilidade, uma tarefa importante é a extração de fluxos de deslocamentos, 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 importante para o mercado e o setor público [Iqbal et al. 2014, Guo et al. 2012].

3.1. Fundamentação

Para se alcançar o objetivo deste trabalho, é preciso que o fluxo de deslocamentos seja extraído de grandes volumes de dados. Para isso, foi desenvolvida uma ferramenta no ambiente Spark³ que é eficiente em termos de processamento computacional, principalmente devido ao uso de paralelismo. Em relação à linguagem, optou-se pelo uso da linguagem Scala, devido aos ganhos de desempenho e expressividade.

A ferramenta proposta pode ser separada em três partes: entrada dos dados, pré-processamento e funcionalidades. Na etapa de entrada, é recebido um conjunto de dados geolocalizados contendo latitude, longitude e algum atributo indicador de tempo, podendo ser por exemplo uma data ou hora. Estes dados devem ser inseridos através de um arquivo de entrada reconhecido pelo Spark, como CSV, Parquet, XML ou Json.

Já na etapa de pré-processamento, o objetivo é preparar os dados de entrada e transformá-los, para a utilização do algoritmo de extração de fluxos. 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 permanência 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*).

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

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

³<https://spark.apache.org/>

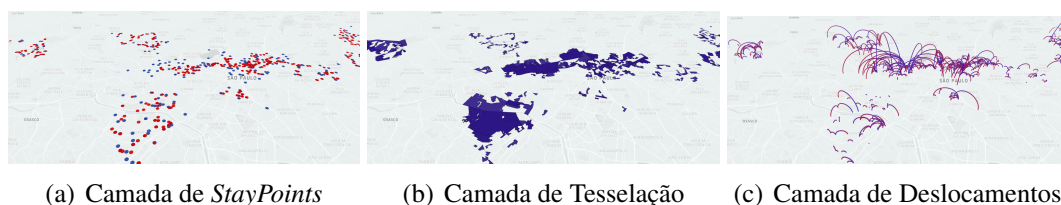


Figura 1. Exemplo dos dados em cada camada

Com os dados preparados, a próxima etapa é a identificação dos fluxos. Para isso, os pontos em instantes do tempo devem ser organizados, inicialmente, nos chamados *StayPoints*. Esses são locais de permanência identificados de acordo com o algoritmo descrito em [Montoliu et al. 2011]. A utilização do algoritmo de *StayPoints* se torna vantajosa, pois filtra, caso os dados sejam densos, a quantidade de pontos intermediários que não representam o início e o fim dos deslocamentos. Podemos observar que a Figura 1 (a) ilustra a identificação de *StayPoints* a partir dos pontos geolocalizados passados como entrada. Os *StayPoints* estão marcados em vermelho na imagem.

Em seguida, cada *StayPoint* deve ser mapeado em uma tesselação. Uma tesselação pode ser definida como um particionamento do espaço, sendo cada partição correspondente a uma região. A tesselação define uma abstração de como fazer a conversão entre pontos e regiões, e pode ser implementada de formas diversas. Neste trabalho, por exemplo, usamos os polígonos representando os setores censitários disponibilizados pelo IBGE. Podemos observar que a Figura 1 (b) ilustra a aplicação das tesselações para identificarmos as regiões a partir dos *StayPoints* passados como entrada. Cada região está marcada de azul na imagem.

Por fim, é feita a identificação dos *Deslocamentos Individuais*. Assume-se que ocorre um deslocamento quando há dois registros do mesmo usuário em regiões diferentes. A Figura 1 (c) ilustra os deslocamentos entre as regiões. Cada linha representa um deslocamento entre um ponto de origem até um ponto de destino, sendo que a origem consiste na cor vermelha e o destino corresponde à cor azul na imagem.

3.2. Extração do Fluxo

Fluxo pode ser definido como a quantidade 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.

Com base nos deslocamentos individuais entre as regiões, é possível calcular o fluxo entre elas. Para esse cálculo, é possível definir uma divisão do tempo de acordo com a necessidade. O horário do dia pode ser categorizado em intervalos de tempo, além dos dias de semana poderem também ser utilizados como categorias. Essas faixas e categorias podem ser definidas pelo usuário da ferramenta no momento da execução.

A saída da extração de fluxo consiste em duas perspectivas: uma matriz origem-destino e uma matriz de regiões. A matriz de origem-destino inclui a quantidade agregada de deslocamentos para cada par de regiões de origem-destino. Essa quantidade é agregada por divisão do tempo, caso a mesma seja definida pelo usuário.

Já a matriz de regiões contém características de cada região individualmente:

- inflow (F_{in}): a quantidade de deslocamentos que chegam na região;
- outflow (F_{out}): a quantidade de deslocamentos que saem na região;
- total flow ($F_{in} + F_{out}$): a quantidade total de deslocamentos, que chegam e saem da região;
- net flow ($F_{in} - F_{out}$): a quantidade líquida de deslocamentos. Se este valor está positivo, indica que um determinado local recebe mais fluxo do que sai; caso esteja negativo, indica que o fluxo de saída deste lugar é maior que o de entrada.

3.3. Análise de Desempenho

Para verificar a eficiência da ferramenta de extração de fluxos proposta, foi feita uma comparação da mesma com o Scikit-Mobility. A Figura 2 mostra os resultados comparando as soluções, sendo que o paralelismo da nossa proposta foi testado para um número variando de 1 a 16 *threads* em paralelo. Está claro que a proposta de extração de fluxo deste trabalho conseguiu diminuir o tempo de processamento de aproximadamente 15 minutos para o Scikit-Mobility, para apenas 2 minutos quando 8 ou 16 *threads* são alocadas. Isso mostra a eficiência da nossa proposta.

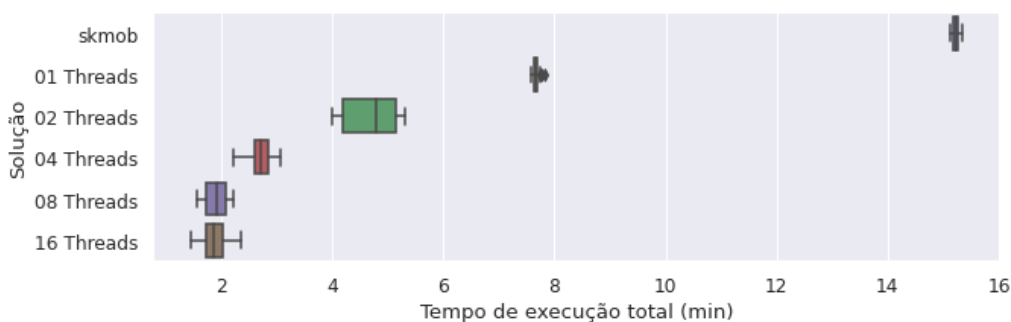


Figura 2. Tempo de processamento para extração de fluxo

4. Análise Centrada no Usuário das Rotas Seguras para Ônibus

Nesta seção é apresentada a avaliação do impacto de se utilizar uma solução de rotas seguras, mais especificamente o trabalho de [Vinícius G. J. Almeida 2022], no deslocamento de usuários de ônibus. Para isso, dados reais de milhares de usuários foram utilizados para que os fluxos sejam extraídos.

4.1. Contextualização

Uma cidade, em geral, possui diversas linhas de ônibus distintas, sendo que cada uma delas é responsável por cobrir determinadas regiões da cidade. Cada linha possui um número de pontos de parada, que são locais obrigatórios e sequenciais por onde o ônibus daquela linha deve passar para embarque e desembarque de passageiros. Entre dois pontos de parada, o ônibus pode seguir diferentes rotas, como a mais curta, a mais segura ou a mais rápida. Além dos pontos de parada, cada esquina por onde o ônibus precisa passar é também considerado um ponto neste trabalho, de acordo com a malha viária da cidade.

Tradicionalmente, entre dois pontos de parada o ônibus tende a seguir pela rota mais curta, isto é, aquela que, considerando as limitações das vias da cidade, percorre a menor distância entre seus pontos de origem e destino (i.e., pontos de parada na

sequência). 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. Essa ideia foi implementada no trabalho [Vinícius G. J. Almeida 2022].

4.2. Construção de Rotas Seguras

O trabalho [Vinícius G. J. Almeida 2022] possui um esquema flexível para o agrupamento de regiões criminais e uma função eficiente para o cálculo do índice de segurança de uma rota. Com base em dados reais de criminalidade disponíveis, é feito um agrupamento das ocorrências e é calculado para cada *cluster* um índice de segurança. Esse índice é utilizado no cálculo de rota segura. Ou seja, se uma rota passa por uma região com alto índice, ela é considerada insegura.

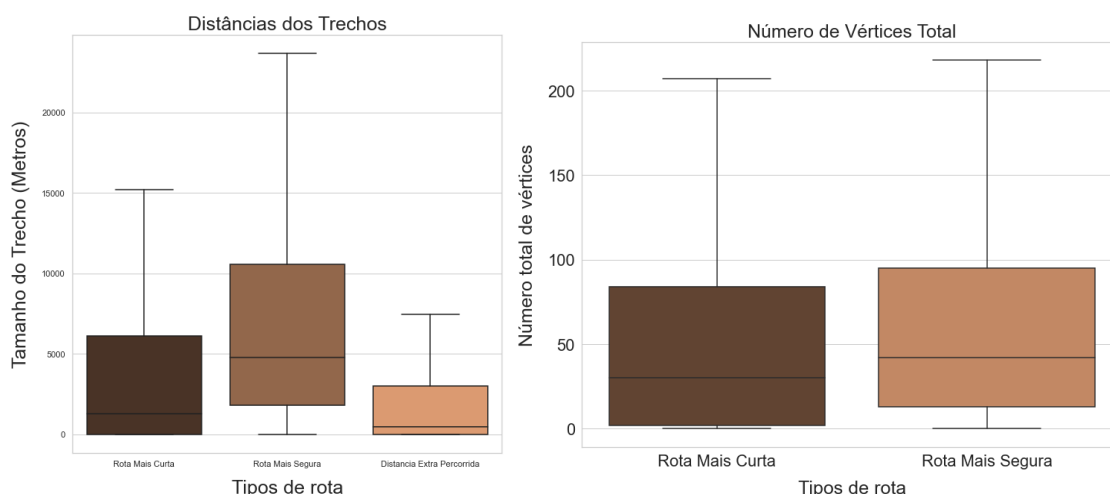
Além de sugerir caminhos que evitam locais com altos índices de criminalidade, a solução proposta ainda permite alterar determinados pontos de paradas presentes em zonas com alta periculosidade para outros mais próximos e menos perigosos. Um ponto de parada só é removido se existe um outro, de qualquer outra linha, fora de uma região criminal e localizado em um raio de até 250m. Isso evita que investimento em infraestrutura para os pontos de parada (i.e., cobertura, painéis de monitoração, etc) seja necessário. Em seguida, as rotas mais curtas e mais seguras são geradas para cada linha de ônibus.

Ao mudar uma rota de uma determinada linha de ônibus, ou realocar algum de seus pontos de parada, os passageiros podem ser afetados pela distância a ser percorrida, já que a rota mais segura nem sempre é a mais curta. Se calcularmos a diferença de comprimento entre a rota mais segura e a mais curta, considerando a linha completa (i.e., do ponto de início ao final), teremos uma visão do impacto na linha como um todo. Mas isso não necessariamente reflete no impacto para os passageiros, pois cada um deles tem pontos de embarque e desembarque diferentes, que raramente coincidem com os pontos inicial e final das linhas. Em outras palavras, 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. Os fluxos de deslocamentos serão utilizados para identificar esses trechos.

4.3. Extração dos Fluxos e Trechos

Para obter os resultados, foram utilizados dados reais de geolocalização, fornecidos por uma empresa parceira sob acordo de confidencialidade. Foram disponibilizados dados de 356.725 usuários, que geraram um total de 11.351.545 registros ao longo de 6 meses do ano de 2021. Boa parte dos usuários contém por volta de 200 registros, o que é uma quantidade considerada razoável para se extrair algum conhecimento.

Para encontrarmos quais trechos das linhas são mais relevantes, utilizamos a ferramenta descrita na Seção 3. Inicialmente, os pontos de cada usuário são agregados em *StayPoints*, que em seguida são mapeados em tesselações. Foram utilizados os polígonos representando cada setor censitário de acordo com o IBGE como regiões. Por fim, são extraídos os fluxos entre cada região a partir dos deslocamentos individuais. Os fluxos foram calculados sem nenhuma categorização ou divisão de tempo, já que para o objetivo do trabalho, o principal é contabilizar todos os deslocamentos, independente do período.



(a) Distribuição da distância dos trechos (b) Distribuição do número de vértices total dos trechos

Figura 4. Tamanho dos trechos e número total de vértices. Os outliers foram removidos para melhor visualização.

Medidas	Vértices Diferentes	Pontos de Parada Diferentes	Vértices Iguais	Pontos de Parada Iguais
Média	15,49	3,76	45,75	6,39
Mínimo	0,00	0,00	0,00	0,00
Máximo	547,00	83,00	493,00	90,00
1o quartil	0,00	0,00	1,00	0,00
Mediana	4,00	0,00	24,00	1,00
3o quartil	13,00	2,00	66,00	9,00
Desvio Padrão	35,72	8,65	59,51	10,45

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

Podemos observar a distância dos trechos na Figura 4(a). Nesta figura, os *outliers* foram removidos para uma melhor visualização. As rotas seguras são maiores que as rotas mais curtas, sendo que a mediana das rotas curtas está bem abaixo dos 5km e próxima de 5km para as rotas mais seguras. Até mesmo no tamanho máximo, podemos observar este aumento significativo em alguns trechos. 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, a Figura 4(b) mostra que o aumento de vértices não chega a ser tão significativo quanto a distância.

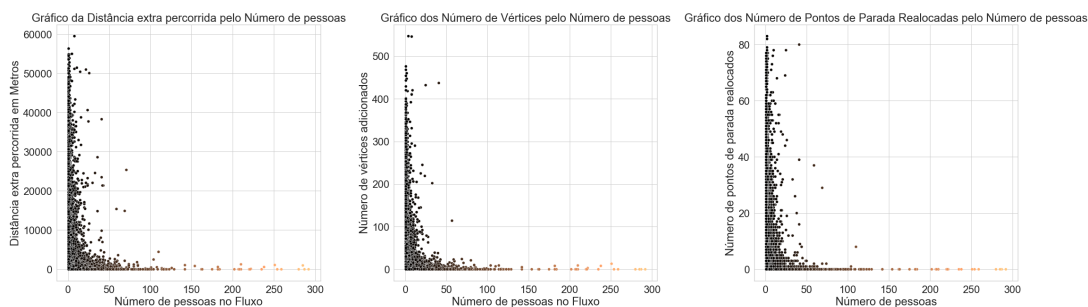
4.4.2. Vértices e Pontos de Parada

Com a mudança da rota de trechos e também com a realocação de pontos de parada, vértices podem ser adicionados e/ou retirados. Para medir o impacto de tais alterações, foi comparado o número de pontos de paradas e vértices da rota mais curta e da mais segura para cada trecho, como pode ser visto na Tabela 1.

Pode-se observar que, em média são alterados poucos vértices entre a rota mais curta e a mais segura, sendo que a grande parte se mantém ou apresenta pouca modificação. Podemos observar também que, em determinados trechos, não há alteração nenhuma, visto que a rota mais curta já é a mais segura. Para 75% dos trechos e para 50% dos trechos, a mudança de vértices é de no máximo 13 e 4, respectivamente. Já em relação aos pontos de paradas, vemos que grande parte das rotas não alteram pontos de parada, ou seja, mantém a integridade da rota original, não afetando assim os passageiros.

4.4.3. Número de Usuários Afetados

Para entendermos o número de pessoas impactadas pelas mudanças nas rotas, podemos olhar pelo fluxo, sendo que cada fluxo corresponde à agregação de deslocamentos individuais. Considerando que cada deslocamento corresponde a uma pessoa, sendo assim, teremos o número total de pessoas.



(a) Distância percorrida pelo número de pessoas (b) Número de vértices no trecho pelo número de pessoas (c) Número de vértices realocados pelo número de pessoas

Figura 5. Número de pessoas afetadas

A Figura 5 (a) apresenta a distância extra percorrida por trecho em relação ao número de pessoas naquele trecho. Pode-se perceber que os trechos com maior distância extra afetam um número pequeno de pessoas, enquanto a maioria dos usuários são pouco impactados. Vale destacar que essa figura contempla todos os trechos, e que alguns deles foram muito afetados quando a rota mais segura é usada, passando a ter um comprimento até 50km maiores. Porém, eles representam um número reduzido.

Já na Figura 5 (b), temos o resultado da análise em relação ao número de vértices adicionados na rota pelo número de pessoas. Por último, na Figura 5 (c) temos a observação do número de pessoas que foram afetadas pelas realocações, e o número de pontos de parada que foram realocados no trecho. Podemos observar que grande parte das mudanças significativas impactam menos de 20 pessoas no trecho, enquanto os de maiores impactos afetam uma parcela bem menor da população.

5. Conclusão e Trabalhos Futuros

Este trabalho apresentou uma análise centrada no usuário da aplicação de uma solução de rotas seguras no transporte coletivo de ônibus na cidade de São Paulo. Foi mostrado que, ao mudar uma rota de um trecho para que ela fique mais segura, o seu comprimento

umenta. Porém, o impacto nos passageiros em si não é muito grande, sendo que a maioria dos trechos utilizados não aumentam mais do que 1 km. Se considerarmos os benefícios de se ter uma rota mais segura, essa distância é aceitável.

Para que essa análise fosse feita, foi desenvolvida uma ferramenta para extração de fluxos a partir de grandes volumes de dados. Essa ferramenta conseguiu reduzir o tempo de processamento de 15 minutos de uma solução existente, para 2 minutos quando 8 processos em paralelo são usados.

Sobre trabalhos futuros, podemos listar a necessidade de se incrementar a ferramenta implementada com novas funcionalidades em relação a fluxos de mobilidade, ou facilitar o cálculo com novas métricas. Também é esperada a realização de outros estudos de caso levando em consideração outras cidades.

Referências

- de Souza, A. and Villas, L. (2020). Vem tranquilo: Rotas eficientes baseado na dinâmica urbana futura com deep learning e computação de borda. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 351–364, Porto Alegre, RS, Brasil. SBC.
- 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*.
- Ladeira, L., de Souza, A., Silva, T., Filho, G. R., Peixoto, M. M., and Villas, L. (2020). Cerva: Roteamento contextual para veículos com risco espaço-temporal. In *Anais do XXXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 379–392, Porto Alegre, RS, Brasil. SBC.
- Ladeira, L., Souza, A., Pereira, G., Silva, T. H., and Villas, L. (2019). Serviço de sugestão de rotas seguras para veículos. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 608–621, Porto Alegre, RS, Brasil. SBC.
- 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.
- Santos, F. A., Rodrigues, D. O., Silva, T. H., Loureiro, A. A. F., and Villas, L. A. (2017). Rotas veiculares cientes de contexto: Arcabouço e análise usando dados oficiais e sensoriados por usuários sobre crimes. In *Anais do XXII Workshop de Gerência e Operação de Redes e Serviços*, Porto Alegre, RS, Brasil. SBC.
- Vinícius G. J. Almeida, Thais R. M. Braga Silva, F. A. S. (2022). Se for, vá na paz: Construindo rotas seguras para veículos coletivos urbanos. In *Anais do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Porto Alegre, RS, Brasil. SBC.