

# Além de um Único Retrato: Análise Temporal da Exposição de Usuários em Dados de Mobilidade

Lucas G. S. Félix<sup>1 2</sup>, Vinícius da F. Vieira<sup>3</sup>,  
Anne Josiane Kouam<sup>2</sup>, Alexandre G. Evsukoff<sup>4</sup>,  
Nadjib Achir<sup>2</sup>, Aline C. Viana<sup>2</sup>, Jussara M. Almeida<sup>1</sup>

<sup>1</sup>Universidade Federal de Minas Gerais (UFMG),

<sup>2</sup>Inria, França,

<sup>3</sup>Universidade Federal de São João del Rei (UFSJ),

<sup>4</sup>Universidade Federal do Rio de Janeiro (UFRJ)

{lucas.felix, jussara}@dcc.ufmg.br

{nadjib, anne-josiane.kouam-djuigne, aline.viana}@inria.fr

vinicius@ufs.j.edu.br

alexandre.evsukoff@coc.ufrj.br

**Abstract.** *The increasing availability of large-scale urban mobility data intensifies privacy risks, as users can often be re-identified from a small number of spatio-temporal observations. Although mobility profiles have been widely used to characterize users' movement behavior, their role in explaining and guiding privacy protection strategies remains unclear, particularly from a longitudinal perspective. In this paper, we investigate the temporal relationship between mobility profiles and user vulnerability in urban mobility data. Our methodology combines mobility profile identification with two complementary vulnerability quantification techniques: uniqueness, capturing spatio-temporal vulnerability, and a hypercube-based approach, capturing behavioral vulnerability. Using mobile phone data from the city of Rio de Janeiro over three consecutive weeks, we conduct a longitudinal analysis of how mobility profiles and vulnerability levels evolve over time. Our results show that, while users frequently change mobility profiles across weeks, their vulnerability levels remain largely stable. Moreover, we find that mobility profiles do not significantly differentiate users' vulnerability, suggesting that profile-based protection strategies may be ineffective. Finally, we show that behavioral vulnerability is consistently concentrated in a small subset of mobility metrics, indicating that protecting these dimensions alone may substantially reduce users' exposure. These findings provide practical insights for the design of targeted and efficient privacy-preserving mechanisms for urban mobility data.*

**Resumo.** *A crescente disponibilidade de dados de mobilidade urbana em larga escala intensifica os riscos à privacidade, uma vez que usuários podem ser re-identificados a partir de um pequeno número de observações espaço-temporais. Embora perfis de mobilidade sejam amplamente utilizados para caracterizar o comportamento de deslocamento dos usuários, seu papel na explicação e no direcionamento de estratégias de proteção à privacidade ainda é pouco compreendido, especialmente sob uma perspectiva longitudinal. Neste trabalho,*

*investigamos a relação temporal entre perfis de mobilidade e a vulnerabilidade dos usuários em dados de mobilidade urbana. Nossa metodologia combina a identificação de perfis de mobilidade com duas técnicas complementares de quantificação de vulnerabilidade: uniqueness, que captura a vulnerabilidade espaço-temporal, e uma abordagem baseada na técnica do hipercubo, que captura a vulnerabilidade comportamental. Utilizando dados de telefonia móvel da cidade do Rio de Janeiro ao longo de três semanas consecutivas, realizamos uma análise longitudinal da evolução dos perfis de mobilidade e dos níveis de vulnerabilidade. Os resultados mostram que, embora os usuários mudem frequentemente de perfil de mobilidade ao longo do tempo, seus níveis de vulnerabilidade permanecem amplamente estáveis. Além disso, observamos que os perfis de mobilidade não diferenciam de forma significativa a vulnerabilidade dos usuários, indicando que estratégias de proteção baseadas em perfis podem ser pouco eficazes. Por fim, mostramos que a vulnerabilidade comportamental se concentra de forma consistente em um pequeno conjunto de métricas de mobilidade, sugerindo que a proteção direcionada a essas dimensões pode reduzir substancialmente a exposição da maioria dos usuários.*

## 1. Introdução

A computação móvel e as telecomunicações tornaram as fronteiras entre a vida real e a digital cada vez mais tênues. Além de remodelarem as dimensões espaciais e temporais da vida social urbana [Ling and Yttri 1999, Frissen 2018], essas tecnologias criaram oportunidades sem precedentes para a coleta e a análise de dados de mobilidade humana, viabilizando avanços em diferentes domínios, como planejamento urbano, monitoramento de epidemias e formulação de políticas públicas [Chaves et al. 2023].

Apesar dessa evolução tecnológica, práticas sociais baseadas no tempo, como horários de trabalho e rotinas familiares, permanecem relativamente estáveis no cotidiano dos indivíduos, mesmo sob forte mediação tecnológica [Green 2002]. Essas rotinas persistentes deixam rastros regulares nos dados de mobilidade, funcionando como verdadeiras “impressões digitais comportamentais”. Embora tais padrões possibilitem a caracterização do comportamento dos usuários e apoiem aplicações relevantes [Amichi et al. 2020], eles também ampliam riscos à privacidade, uma vez que dados individuais de mobilidade são altamente distintivos e propensos a reidentificação [De Montjoye et al. 2013, Farzanehfar et al. 2021, Félix et al. 2025].

No contexto da mobilidade urbana, a vulnerabilidade dos usuários pode se manifestar de diferentes formas. Neste trabalho, consideramos dois tipos principais de vulnerabilidade: (i) A **vulnerabilidade espaço-temporal** explora padrões explícitos de movimentação, analisando trajetórias cruas de usuários no espaço e no tempo [Zang and Bolot 2011, De Montjoye et al. 2013, Farzanehfar et al. 2021], e a (ii) **vulnerabilidade comportamental**, associada a padrões implícitos de comportamento, revelados por meio de representações derivadas dos dados de mobilidade, como *embeddings* ou métricas agregadas que capturam hábitos individuais [May Petry et al. 2020, Félix et al. 2025]. Enquanto a primeira está diretamente ligada à singularidade das trajetórias, a segunda está ligada à singularidade comportamental. Em ambos os casos, a vulnerabilidade pode ser estimada individualmente.

Neste cenário, a aplicação de técnicas de proteção à privacidade, como *k-anonymity* e *differential privacy*, permite a proteção dos usuários, ao aplicar a todos

os usuários técnicas de introdução de ruídos nos dados que fazem com que estes sejam indistinguíveis. No entanto, essas técnicas apresentam custos computacionais elevados e envolvem um *trade-off* entre privacidade e utilidade dos dados [Jiang et al. 2013, Gramaglia et al. 2021]. Assim, a proteção personalizada, i.e., guiada por técnicas que quantificam a vulnerabilidade do usuário, surge como uma possível técnica para aumentar a utilidade e proteção dos dados. Porém, em grandes conjuntos de dados de mobilidade urbana, como aqueles provenientes de operadoras telefônicas, proteger todos os usuários de forma personalizada é, na prática, inviável. Nesse contexto, identificar grupos de usuários com características e comportamentos semelhantes surge como uma alternativa promissora para orientar estratégias de proteção mais eficientes.

Seguindo essa linha, trabalhos anteriores propuseram a caracterização de perfis de mobilidade que sintetizam padrões espaço-temporais recorrentes dos usuários. Propostas como a de [Amichi et al. 2020] identificam três perfis principais: **Rotineiros**, que visitam frequentemente os mesmos locais; **Exploradores**, que tendem a buscar novidades e variar seus deslocamentos; e **Regulares**, que apresentam um equilíbrio entre rotina e exploração. Esses perfis têm sido amplamente utilizados para descrever comportamentos de mobilidade, mas seu potencial para orientar mecanismos de preservação de privacidade ainda é pouco explorado.

Apesar dos avanços na análise de perfis de mobilidade e no estudo da vulnerabilidade de usuários, esses dois conceitos, perfil-vulnerabilidade, nunca foram avaliados sob uma mesma ótica. Não obstante, as análises feitas sob perspectiva da vulnerabilidade são no geral estáticas, avaliando apenas períodos finitos de tempo, o que é, no geral, uma prática comum devido a limitações nos dados. Neste cenário, ainda não está claro se usuários com padrões de mobilidade semelhantes apresentam níveis similares de vulnerabilidade, nem como essa relação evolui ao longo do tempo e em diferentes contextos urbanos. Essa ausência de uma perspectiva temporal limita o desenvolvimento de estratégias de proteção mais eficazes, que podem ser personalizadas de maneira a considerar diferenças sistemáticas de risco entre grupos de usuários.

Diante desse cenário, nossa proposta visa **expandir as avaliações feitas sob técnicas que quantificam a vulnerabilidade**, avaliando como a relação perfil-vulnerabilidade se comporta em um cenário longitudinal, i.e., como uma avaliação temporal mais detalhada. Partimos da hipótese de que usuários com perfis de mobilidade semelhantes tendem a compartilhar padrões de vulnerabilidade comportamental e espaço-temporal, o que pode ser explorado para apoiar mecanismos de preservação de privacidade mais direcionados. Mais especificamente, este trabalho busca responder às seguintes questões de pesquisa:

- **RQ1:** Quais são as diferenças de vulnerabilidade entre perfis?
- **RQ2:** Como a relação entre perfil de mobilidade e vulnerabilidade evolui ao longo do tempo?

Para responder a essas questões, identificamos perfis de mobilidade utilizando a metodologia proposta por [Amichi et al. 2020] e quantificamos a vulnerabilidade individual dos usuários por meio de duas abordagens complementares: A vulnerabilidade espaço-temporal é estimada pela métrica de *Uniqueness* [De Montjoye et al. 2013], enquanto a vulnerabilidade comportamental é capturada pela abordagem de Hipercubo [Félix et al. 2025], que foi previamente proposta por nós. Note que nesse trabalho nós não propomos uma nova técnica para quantificar a vulnerabilidade, mas expandimos a compreensão da vulnerabilidade através da análise de perfis e temporal.

Realizamos a avaliação deste trabalho em conjunto com dados de telefonia da cidade do Rio de Janeiro, o qual compreende três semanas de dados. Dadas nossas questões de pesquisa, este conjunto de dados é ideal devido à sua maior escala temporal, a qual permite uma análise individual e temporal dos indivíduos. As principais contribuições deste trabalho são:

- Uma análise longitudinal inédita da relação entre perfis de mobilidade e vulnerabilidade, investigando como ambos evoluem ao longo do tempo em dados reais de mobilidade urbana.
- Uma evidência empírica de que perfis de mobilidade não são bons discriminadores de vulnerabilidade, uma vez que usuários com perfis distintos apresentam níveis de exposição semelhantes, mesmo em diferentes janelas temporais.
- A demonstração de que a vulnerabilidade individual tende a ser estável ao longo do tempo, apesar de mudanças frequentes nos perfis de mobilidade dos usuários, indicando que o risco de exposição é uma característica persistente do comportamento individual.
- A identificação de um pequeno conjunto de métricas comportamentais que concentra a maior parte da vulnerabilidade, independentemente do perfil do usuário, sugerindo que estratégias de proteção direcionadas a essas dimensões podem ser suficientes para mitigar a exposição da maioria dos usuários.

O restante do artigo está organizado da seguinte forma. A Seção 2 apresenta a contextualização e os trabalhos relacionados. A Seção 3 descreve as técnicas utilizadas para identificação de perfis e quantificação de vulnerabilidade. O conjunto de dados é apresentado na Seção 4. A Seção 5 apresenta resultados das análises e responde às questões de pesquisa. Por fim, a Seção 6 apresenta as conclusões.

## 2. Trabalhos Relacionados

A crescente disponibilidade de dados de mobilidade através de celulares e aplicações tem permitido que empresas projetem e avaliem serviços orientados por dados de mobilidade, resultando em um ecossistema avaliado em bilhões de dólares [Analytica 2025]. Contudo, devido ao fato desses dados carregarem informações sensíveis, que podem expor os usuários presentes nesses conjuntos de dados, estes não são facilmente compartilhados. Isso é corroborado por diversos trabalhos na literatura que mostram que apenas alguns pontos espaço-temporais são suficientes para reidentificar mais de 90% dos usuários de um conjunto de dados de mobilidade [Zang and Bolot 2011, De Montjoye et al. 2013]. Se considerarmos a adição de informações externas, como o perfil de mobilidade do usuário [Amichi et al. 2020], a taxa de reidentificação pode ficar ainda mais alta [May Petry et al. 2020].

Neste cenário, técnicas para a redução do nível de vulnerabilidade em conjuntos de dados como *k-anonymity* e *differential privacy*, que protegem os usuários de maneira uniforme, auxiliam na redução da vulnerabilidade individual. Contudo, essas técnicas frequentemente implicam alto custo computacional e significativa perda de utilidade dos dados [Jiang et al. 2013]. Nesse contexto, técnicas de quantificação da vulnerabilidade podem auxiliar tais mecanismos ao permitir a identificação e a proteção direcionada apenas dos usuários mais vulneráveis.

Como discutido previamente, as técnicas de quantificação de vulnerabilidade podem ser classificadas em duas categorias principais: vulnerabilidade

espaço-temporal e vulnerabilidade comportamental. Abordagens como a *uniqueness* [De Montjoye et al. 2013] quantificam a vulnerabilidade espaço-temporal ao analisar visitas, ou combinações de visitas temporalmente ordenadas, realizadas pelos usuários, identificando aquelas que são únicas. Assim, um usuário é considerado vulnerável caso apresente ao menos uma visita única, i.e., não é feita por mais nenhum usuário do conjunto de dados durante aquela janela temporal. Por outro lado, a vulnerabilidade comportamental, como abordada em nosso trabalho previamente publicado no SBRC 2025 [Félix et al. 2025], busca quantificar a vulnerabilidade intrínseca dos usuários a partir de métricas de mobilidade em espaços de alta dimensionalidade. Contudo, devido às restrições de compartilhamento de dados, grande parte das avaliações em mobilidade urbana permanece limitada a análises estáticas, isto é, não consideram a evolução temporal dos padrões de mobilidade dos usuários [Félix et al. 2025, Pyrgelis et al. 2018]. Em particular, técnicas baseadas no comportamento do usuário [May Petry et al. 2020, Félix et al. 2025], que dependem de períodos de observação mais longos para que as métricas se estabilizem [Liu et al. 2018, Su et al. 2022], são, em geral, avaliadas sob uma perspectiva estática.

À luz desses fatos, o objetivo deste trabalho é estender a avaliação da vulnerabilidade em dados de mobilidade ao incorporar explicitamente a relação entre perfil de mobilidade e vulnerabilidade, bem como analisar como essa relação evolui ao longo do tempo. Em vez de propor novas técnicas, este estudo se baseia em métodos já consolidados na literatura, aplicando-os de forma sistemática para investigar a dinâmica temporal da relação perfil–vulnerabilidade. Assim, nós estendemos a avaliação de nossa proposta feita previamente [Félix et al. 2025], utilizando uma metodologia complementar para a avaliação de perfis. Além disso, utilizamos outro conjunto de dados, visto que os dados previamente utilizados capturavam um período temporal reduzido comparado ao que utilizamos aqui. Assim, a **principal contribuição reside em uma avaliação longitudinal dessa relação perfil-vulnerabilidade, conduzida sob as perspectivas de técnicas de quantificação da vulnerabilidade espaço-temporal e comportamental.**

### 3. Metodologia

Nesta seção, apresentamos a metodologia utilizada para a condução da investigação proposta na direção de uma compreensão das Questões de Pesquisa levantadas na Seção 1, buscando estender a avaliação de técnicas para quantificação de vulnerabilidade através do estudo da relação perfil-mobilidade e realizando uma avaliação longitudinal dessa relação. Para isso, utilizamos uma metodologia dividida em duas etapas principais: (i) Primeira etapa focada em identificar os perfis de mobilidade dos usuários, (ii) Segunda etapa focada na quantificação das vulnerabilidades dos usuários. Abaixo, nós apresentamos como cada uma das técnicas foi desenvolvida e referenciamos os artigos originais para maiores detalhes.

#### 3.1. Identificação de perfis de mobilidade:

Para identificação dos perfis de mobilidade de usuários, empregamos a proposta de [Amichi et al. 2020]. A proposta identifica três principais perfis de mobilidade observados em conjuntos de dados de mobilidade:

- **Exploradores:** Usuários que frequentemente exploram novas áreas, visitando lugares nunca visitados anteriormente.
- **Rotineiros:** Usuários que geralmente permanecem em uma rotina, em torno do eixo casa–trabalho, embora apresentem momentos de novidade.

- **Regulares:** Usuários que possuem um comportamento mais equilibrado entre rotina e momentos de busca por novidade.

A metodologia proposta por [Amichi et al. 2020] para identificação dos perfis é dividida em duas etapas: (i) primeiramente, cada local visitado por um usuário é categorizado de acordo com sua frequência de visita. Os locais são caracterizados como exploração (i.e., um local novo ou raramente visitado) ou retorno (i.e., um local frequentemente visitado). Uma vez que os locais são classificados, duas métricas são investigadas: **Intermitência**, que é a frequência com que os usuários alternam entre explorações e retornos, e **Grau de Retorno**, isto é, a frequência de revisitas a um local do tipo retorno após uma visita a um do tipo exploração. (ii) Uma vez obtidas essas características para cada usuário, elas são agrupadas utilizando a técnica *K-Means*, sendo o parâmetro  $k$ , que define o número de grupos, igual a três. Avaliamos a qualidade dos grupos gerados utilizando a técnica da silhueta.

### 3.2. Quantificação da vulnerabilidade de usuários:

Após a definição de cada perfil, nós quantificamos o nível de vulnerabilidade de cada um dos indivíduos utilizando duas técnicas: *Uniqueness* [De Montjoye et al. 2013] e *Hipercubo* [Félix et al. 2025].

#### Uniqueness

A *Uniqueness* mede a exposição dos usuários ao quantificar a singularidade na combinação de visitas [De Montjoye et al. 2013] em diferentes janelas temporais de agregação ( $t_w$ ), isto é, quanto mais exclusivo for o padrão de mobilidade do usuário durante um período de tempo, mais exposto este usuário estará.

#### Hipercubo

O Hipercubo, proposto previamente no trabalho [Félix et al. 2025], quantifica a vulnerabilidade comportamental de usuários explorando métricas de mobilidade urbana (e.g., raio de giro, diversidade, regularidade), caracterizando a vulnerabilidade de cada usuário pela quantidade dos outros usuários que compartilham comportamento semelhante. Os usuários são considerados vulneráveis se tiverem um valor de  $uniqueness_i = 1$ , i.e., não dividem a trajetória com mais ninguém dentro daquela janela de tempo.

Para modelar o comportamento dos usuários, o Hipercubo extrai de cada usuário  $i \in U$ , onde  $U$  é o conjunto de usuários em um conjunto de dados de mobilidade, um conjunto de métricas  $\mathcal{M} = \{m_1^i, m_2^i, \dots, m_r^i, \dots, m_m^i\} \in \mathbb{R}^+$ , onde cada componente  $m_r^i$  representa o valor da métrica  $m_r$  para o usuário  $i$ . Neste espaço, é construído ao redor de cada  $i$  um Hipercubo de tamanho  $v$ , o qual define uma área de indistinguibilidade entre usuários  $i$  e  $j$ . A construção do Hipercubo pode ser formulada como apresentado na Equação 1, onde uma matriz binária  $I_{i,j}^r$  considera todas as métricas  $r$  em que os usuários  $i$  e  $j$  são vizinhos ( $I_{i,j}^r = 1$ ), ou não ( $I_{i,j}^r = 0$ ). Dois usuários  $i$  e  $j$ , são considerados vizinhos, se e somente se, eles forem vizinhos em todas as dimensões, i.e., o somatório de  $I_{i,j}^r$  para todas as métricas  $r$  deve ser igual ao número total de dimensões  $|M|$  (Equação 2). Por fim, o nível de vulnerabilidade dos usuários é calculado somando a quantidade de vizinhos diferentes dentro do Hipercubo de  $i$ , como mostrado na Equação 3. Assim, um usuário é vulnerável se ele estiver sozinho dentro deste Hipercubo e está protegido se

houver pelo menos um outro usuário dentro deste Hipercubo, visto que não é possível distinguir entre o comportamento destes dentro de uma variação máxima  $v$ .

$$I_{i,j}^r = \begin{cases} 1, & \text{if } (1 - v) \times m_r^i \leq m_r^j \leq (1 + v) \times m_r^i, \forall j \in U, i \neq j \\ 0, & \text{caso contrário} \end{cases} \quad (1)$$

$$B_{i,j} = \begin{cases} 1, & \text{if } \sum_{r=0}^{|M|} I_{i,j}^r = |M|, \forall j \in U, i \neq j \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

$$NC_i = \sum_{j=1, j \neq i}^{|U|} B_{i,j} \quad (3)$$

Vale ressaltar que a técnica do Hipercubo é altamente interpretável, pois permite identificar as métricas que mais contribuem para a vulnerabilidade dos usuários. De acordo com esse método, quando um outro usuário está localizado dentro do Hipercubo, ele pode ser considerado protegido, uma vez que não é possível distingui-lo de seus vizinhos pelo seu comportamento de mobilidade. A vulnerabilidade do usuário pode, então, ser interpretada em função da distância ao vizinho mais próximo: quanto maior essa distância para a dimensão mais afastada, mais distante o usuário se encontra de uma condição de proteção, já que devem ser vizinhos em todas as dimensões para que a condição de vizinhança seja satisfeita. Dessa maneira, é possível definir o vizinho mais próximo de um usuário  $i$  como aquele usuário  $j$  do qual se observa a menor distância em sua dimensão mais distante, o que pode ser calculado pela distância relativa no Hipercubo, dada pela Equação 4. A partir da identificação do vizinho mais próximo, é possível calcular a distância relativa a cada uma das métricas. Por fim, o valor  $P_{i,j}$  nos permite identificar a dimensão na qual que um usuário teria que variar mais para que seu comportamento fique mais próximo ao de seu vizinho  $j$ .

$$P_{i,j} = \max_{r \in M} \left( \frac{|m_r^i - m_r^j|}{m_r^i} \right) \quad (4)$$

É importante observar que, embora essas técnicas compartilhem um objetivo semelhante – a quantificação da exposição do usuário –, elas capturam perspectivas diferentes. *Uniqueness* explora apenas os **padrões espaço-temporais**, observando somente as trajetórias dos usuários. Hipercubo explora os **padrões comportamentais** dos usuários, extraindo métricas de mobilidade (como raio de giro, diversidade, regularidade) de cada trajetória que caracterizam o comportamento de mobilidade do usuário. Assim, usuários que apresentam um comportamento muito distinto de seus pares são considerados propensos à reidentificação. A utilização dessas duas técnicas traz perspectivas complementares à análise da vulnerabilidade de usuários em uma base de dados de mobilidade.

De forma a manter a consistência e permitir a comparação com outros trabalhos na literatura, definimos a mesma configuração experimental adotada em [Félix et al. 2025]:  $t_w = 4$  para a janela de agregação da técnica *uniqueness*, e  $v = 0.1$  (10% de variação máxima no comportamento) para a instanciação do Hipercubo. Ainda de acordo com [Félix et al. 2025], foi definido um espaço  $m$ -dimensional, considerando um grupo de 13

métricas divididas em três grupos: **Espaciais** (Raio de Giro, 2k Raio de Giro, Distância Máxima, Número de Visitas, Média e Desvio Padrão (DP) do Número de Saltos, e Número de Visitas, e Número de Visitas Únicas); **Temporais** (Média e DP do tempo entre visitas); e **Estruturais** (Entropia, Estacionariedade e Regularidade). É importante notar que, as métricas utilizadas para capturar o perfil do usuário, intermitência e grau de retorno, não são utilizadas para o cálculo da vulnerabilidade utilizando o Hipercubo.

#### 4. Estudo de Caso - Conjuntos de Dados

Neste trabalho, estendemos a avaliação de técnicas para quantificação de vulnerabilidade, realizando uma análise perfil-vulnerabilidade de maneira longitudinal. Para atingir tal objetivo, consideramos dados de mobilidade urbana na cidade do Rio de Janeiro-RJ. Nesta seção, detalhamos as etapas de pré-processamento e apresentamos uma breve descrição dos dados após o pré-processamento.

Os dados de mobilidade dos indivíduos considerados neste trabalho são construídos a partir de dados CDR (*Call Detail Records*), registros de ligações telefônicas armazenados pelas operadoras com a finalidade de adequar a cobrança de seus clientes. É importante ressaltar que os dados CDR utilizados neste trabalho foram obtidos junto a uma grande operadora de telefonia, de forma anonimizada e estão restritos a fins acadêmicos do grupo de pesquisa. Os dados CDR armazenam, entre outras informações, os IDs dos indivíduos que efetuaram ligações telefônicas, além do instante e da localização dessa ocorrência. Assim, se um indivíduo efetua ligações telefônicas em duas localidades distintas em um determinado intervalo de tempo, é possível deduzir o seu deslocamento entre essas localidades. Por isso, dados dessa natureza são inerentemente esparsos, uma vez que o registro da localização do usuário depende de sequências de ações ativas, como a realização de ligações ou o envio de mensagens. Assim, apesar da base original conter quase 2 milhões de usuários em todo o estado do Rio de Janeiro, as análises deste trabalho, são restritas ao município do Rio de Janeiro, que concentra a maior densidade de usuários ativos, permitindo maior robustez estatística nas estimativas. Para garantir a reconstrução de trajetórias diárias minimamente informativas, foi selecionado o período consecutivo que maximiza o número de usuários com ao menos cinco registros de localização por dia. Esse limiar representa o mínimo necessário para capturar padrões diários não triviais de mobilidade, conforme indicado por estudos anteriores [Liu et al. 2018, Su et al. 2022], mesmo a base de dados original abrangendo o período de julho a dezembro de 2012. Tal critério implica uma redução significativa da amostra, uma vez que poucos usuários apresentam elevada atividade de comunicação de forma consistente ao longo de dias consecutivos. Adicionalmente, foram selecionadas semanas consecutivas com o objetivo de preservar a continuidade temporal das trajetórias, evitando a introdução de quebras artificiais que poderiam afetar métricas dependentes de rotina e regularidade. Dentre os períodos elegíveis, foram escolhidas três semanas consecutivas que maximizam o número de usuários, totalizando aproximadamente 3840 usuários distintos.

Com o intuito de reduzir vieses associados à esparsidade temporal e espacial dos dados, foi aplicado um *pipeline* de pré-processamento composto por quatro etapas, conforme descrito em [Esper et al. 2024]:

- **Discretização temporal:** Registros duplicados dentro da mesma hora foram removidos, garantindo no máximo uma localização por usuário por hora. Essa

discretização reduz redundâncias e viés na extração de métricas, ao mesmo tempo em que preserva a resolução temporal necessária para caracterizar padrões diários de mobilidade.

- **Tesselação espacial:** Para mitigar a esparsidade espacial, foi aplicada uma grade regular de 200 m × 200 m.
- **Completação de trajetórias:** Lacunas temporais foram reduzidas por meio da inferência de permanência em locais previamente identificados como residência ou trabalho [Amichi et al. 2020, do Couto Teixeira et al. 2021]. Essa etapa é estritamente restrita a locais já observados, não introduzindo novos pontos espaciais nem alterando a diversidade espacial das trajetórias. Dessa forma, busca-se reduzir lacunas temporais, que são comuns em dados de telefonia.

A Tabela 1 apresenta as estatísticas para essas três semanas após o pré-processamento.

**Tabela 1. Estatísticas semanais dos dados após pré-processamento.**

Dataset	# Usuários	# Locais	# Dias	Média de locais por usuário	Média de locais por usuário por dia	Média de locais dif. por usuário	Média de locais dif. por usuário por dia
RJ - #1	3840	859	7	69,7	9,95	9,49	3,31
RJ - #2	3840	886	7	70,13	10,01	9,52	3,12
RJ - #3	3840	848	7	69,21	9,88	9,44	3,08

## 5. Resultados

Nesta seção, nós apresentamos os resultados dessa pesquisa. Inicialmente, nós apresentamos e discutimos como os perfis de usuários se dividem ao longo das semanas e como eles variam. Depois, apresentamos os níveis de vulnerabilidade e focamos na análise de como o perfil-vulnerabilidade se relaciona, endereçando as duas questões de pesquisa.

### 5.1. Análise de perfis de mobilidade

A Tabela 2 apresenta a distribuição percentual dos perfis de mobilidade ao longo das semanas, bem como seus níveis de estabilidade (associada à sustentação do mesmo perfil ao longo das semanas) e instabilidade (associada à mudança de perfil no período) desses perfis. Na tabela é apresentada a estabilidade/instabilidade relativa ao perfil. Note que os valores de estabilidade e instabilidade são complementares.

A partir da análise da distribuição percentual dos perfis ao longo das semanas, observa-se que o conjunto de dados é predominantemente composto por usuários Regulares, seguidos por Exploradores e Rotineiros. A maior representatividade dos usuários Regulares é corroborada pela média diária de locais distintos visitados, em torno de três, geralmente associados a casa, trabalho e um local adicional de exploração. Esse padrão se mantém consistente ao longo das semanas, refletindo a estabilidade observada na distribuição dos perfis de mobilidade.

Não obstante, o padrão dos perfis se repete ao longo das três semanas, com uma variação máxima de 6% na proporção de usuários em cada perfil, o que indica estabilidade na distribuição dos perfis. Contudo, ao analisar a proporção de usuários que permanecem estáveis ou instáveis ao longo das semanas, essa conclusão mostra-se apenas parcialmente válida, uma vez que os níveis de estabilidade e instabilidade dependem do perfil de mobilidade. Usuários de perfil regular possuem uma maior chance de estabilidade no perfil,

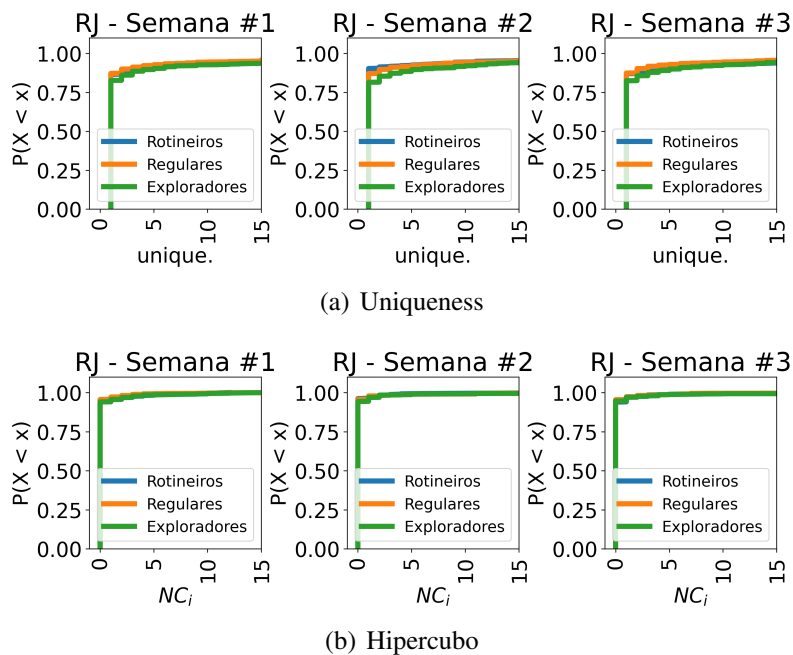
enquanto os perfis exploradores e rotineiros são, no geral, mais instáveis. Vale ressaltar que exploradores e rotineiros correspondem a perfis com características mais extremas, uma vez que tendem, respectivamente, a visitar muitos locais distintos ou um número muito reduzido de locais. Contudo, devido ao padrão bem estabelecido da mobilidade cotidiana, geralmente marcado por uma estrutura estável de deslocamento entre casa e trabalho, esses perfis de alta exploração ou de rotinização extrema mostram-se difíceis de serem mantidos ao longo do tempo.

Esses resultados indicam que há, no geral, uma baixa estabilidade por perfil, indicando uma mudança frequente de perfil, reforçando o caráter dinâmico da mobilidade de usuários, indicando que períodos curtos de observação podem não ser o suficiente para propriamente caracterizá-los, limitando o potencial benefício da utilização de tais perfis em métodos de sistemas de recomendação e mecanismos de proteção de dados.

**Tabela 2. Porcentagem de perfis por semana - Nível de estabilidade/instabilidade**

Perfil	Semana			Total	Estável			Instável		
	1	2	3		1-2	1-3	2-3	1-2	1-3	2-3
Exploradores	34,19%	28,44%	29,06%	30,56%	33,7%	34,23%	30,56%	66,3%	65,77%	69,44%
Regulares	53,78%	58,78%	57,84%	56,8%	53,74%	52,68%	59,34%	46,26%	47,32%	40,66%
Rotineiros	12,03%	12,79%	13,1%	12,64%	11,2%	10,54%	12,92%	88,8%	89,46%	87,08%

## 5.2. Análise perfil-vulnerabilidade



**Figura 1. Métricas de vulnerabilidade por perfil ao longo de semanas**

Agora, direcionamos nossa análise para avaliação perfil-vulnerabilidade. As Figuras 1(a), 1(b) apresentam as distribuições de *Uniqueness* e Hipercubo para cada uma das semanas. Vale lembrar que para ambas as técnicas consideramos o cenário binário da vulnerabilidade: (i) Caso a *Uniqueness* seja igual a 1, significa que ninguém compartilha com o usuário aquela trajetória, e conseqüentemente ele está vulnerável, (ii) Caso o número de vizinhos dentro do Hipercubo seja igual a 0, o usuário está vulnerável, visto

que não há ninguém que possua um comportamento comparável considerando  $v = 0.1$  (i.e. 10% de variação).

Neste cenário, é possível observar que grande parte da população, mais de 75% independente da métrica para cálculo da vulnerabilidade, perfil e semana, é apontada como vulnerável. Observamos que o Hipercubo captura uma porcentagem maior de usuários como vulneráveis que a *uniqueness*, o que indica uma maior vulnerabilidade comportamental, em detrimento do espaço-temporal. Tal diferença é justificada pela alta distinguibilidade do comportamento, como previamente observado [Félix et al. 2025].

Além disso, também independentemente da métrica, é possível perceber que a relação perfil-vulnerabilidade é pouco distinguível, não sendo possível identificar um maior nível de vulnerabilidade dado o perfil do usuário. Neste cenário, uma das vantagens da técnica do Hipercubo é a sua alta interpretabilidade, a qual permite identificarmos se há métricas a que cada perfil se associa mais quando os usuários estão vulneráveis. As Figuras 2 apresentam para cada perfil a porcentagem de usuários vulneráveis em função da dimensão mais difícil de ser protegida por semana, isto é, métricas que se encontram mais distantes do vizinho mais próximo e tornam a proteção do usuário vulnerável ( $NC_i = 0$ ) mais difícil, conforme definido na Equação 4.

A partir dessa análise é possível identificar com maior precisão às relações entre os perfis de usuários em função de sua vulnerabilidade. Primeiramente, o que percebemos é um padrão bem definido para as métricas que mais contribuem e as quais menos contribuem: (i) **Maior contribuição:** # de Locais Distintos, 2k Raio de Giro, Distância Máxima, Média de Saltos e DP Tempo, (ii) **Menor contribuição:** # de Visitas, DP Saltos, Regularidade, Diversidade e Média Tempo. Neste cenário, tanto o ranking das métricas das top-5 métricas que mais contribuem e as top-5 que menos contribuem para a vulnerabilidade possuem um coeficiente de *Jaccard* igual a 1, o que implica que protegendo essas 5 dimensões principais é possível proteger grande parte dos usuários independente de seu perfil.

Neste cenário, o que percebemos é que as métricas de maior contribuição capturam especificidades da mobilidade do usuário que tornam um usuário único como a Distância Máxima, a qual captura a maior distância percorrida pelo usuário, capturando seus padrões atípicos de deslocamento, assim como o # de Locais Distintos, o qual captura o # de locais diferentes que esse usuário visita, 2k Raio de Giro, que irá capturar o quanto esse usuário desloca considerando como pontos centrais casa e trabalho, e DP Tempo, o qual irá capturar a variação nos padrões de comunicação do usuário. Por outro lado, as métricas que contribuem menos para a identificação da vulnerabilidade dos usuários são aquelas fortemente influenciadas por vieses nos padrões de comunicação. Esses vieses fazem com que os usuários apresentem, em geral, valores muito semelhantes para tais métricas, resultando em baixa variabilidade e reduzida capacidade de distinção. Um exemplo disso é a métrica de número de visitas, que depende diretamente da quantidade de ligações realizadas pelo usuário ao longo do dia. Como o conjunto de dados considera apenas usuários que efetuam no mínimo 5 ligações diárias, esses usuários tendem a apresentar números de ligações semelhantes. Consequentemente, essa restrição limita a capacidade de capturar de forma detalhada outras métricas como a diversidade e a regularidade, uma vez que o comportamento do usuário não é observado de maneira contínua ao longo de todo o dia. Contudo, ressaltamos que a utilização de um período longo de observação e a utilização de um conjunto de métricas amplo minimizam esses vieses e permitem capturar a vulnerabilidade dos usuários, uma vez que a técnica do

Hipercubo considera que se o usuário estiver vulnerável em uma dimensão, logo, ele é consequentemente vulnerável em todas dimensões.

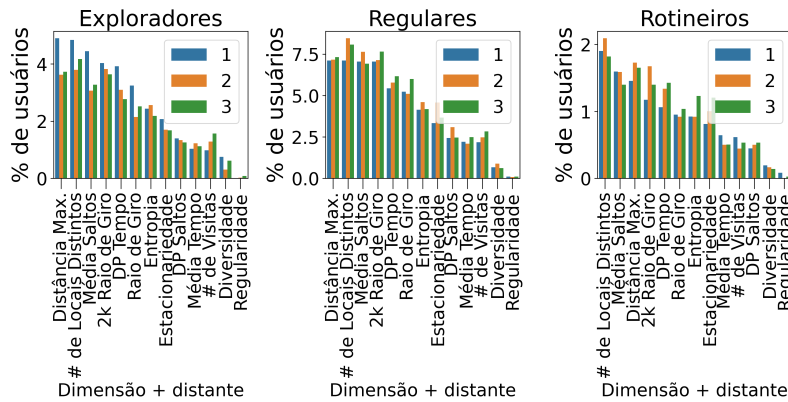


Figura 2. Dimensões mais vulneráveis por perfil

Essa análise nos ajuda a responder à nossa primeira questão de pesquisa (RQ1), a qual visa entender as diferenças entre perfil de mobilidade e a vulnerabilidade do usuário: Há pouca diferença entre a vulnerabilidade observada em diferentes perfis, visto que não conseguimos observar isso nas distribuições de *uniqueness* e Hipercubo por perfil. Além disso, observamos que há pouca variabilidade nas métricas que mais contribuem para a vulnerabilidade de diferentes usuários, o que implica que uma técnica de proteção que consiga proteger essas 5 dimensões, seria capaz de proteger grande parte dos usuários. Em suma, para a RQ1 o que observamos é que não há diferença de vulnerabilidade entre os usuários de acordo com o perfil.

Com relação à segunda questão de pesquisa (RQ2), a qual foca na relação perfil-vulnerabilidade ao longo do tempo, nós discutimos e mostramos previamente que há uma estabilidade nas relações entre perfil-vulnerabilidade ao longo das semanas. Nas análises anteriores foi possível perceber que a porcentagem de usuários capturados pelas técnicas de quantificação de vulnerabilidade se mantém similar. Contudo, como há uma instabilidade no perfil desses usuários, o que mostra que os usuários tendem a mudar a maneira que transitam ao longo do tempo, isso pode indicar que há instabilidade também na vulnerabilidade, i.e., que os usuários se movem entre um estado vulnerável/protegido, assim como se movem entre os perfis. A Tabela 3 nos auxilia em tal avaliação, permitindo analisar a porcentagem de usuários que mudaram de perfil e de estado de vulnerabilidade ao longo das semanas. Neste cenário, ao avaliar a tabela, o que percebemos é que apesar da mudança de perfil em média 90% dos usuários mantêm o mesmo nível de *uniqueness* ao longo da semana, enquanto 95% dos usuários em média mantêm o mesmo número de vizinhos dentro do Hipercubo. Isso nos mostra que, apesar dos usuários mudarem de perfil, seu nível de vulnerabilidade se mantém estável ao longo do tempo.

Com isso, o que nós observamos para a segunda questão de pesquisa RQ2 é que a relação perfil-vulnerabilidade é mutável ao longo do tempo, no sentido de que os usuários tendem a mudar de perfil, porém, mantêm um nível similar de vulnerabilidade ao longo do tempo. Isso indica que por mais que os usuários mudem as características de sua mobilidade, eles ainda podem se manter vulneráveis ao longo das semanas. Tal comportamento nos permite tirar as seguintes conclusões: (i) As relações individuais entre perfil de mobilidade e vulnerabilidade são mutáveis. Embora os usuários tendam a manter níveis semelhantes de vulnerabilidade ao longo do tempo, seus perfis de mobilidade podem va-

riar, o que é coerente com a dinâmica interpessoal e social. Algumas semanas apresentam padrões de mobilidade mais intensos e exploratórios, enquanto outras são mais estáveis e rotineiras. (ii) Usuários apresentam uma tendência de se manterem no mesmo estado de vulnerabilidade, tanto espaço-temporal quanto comportamental; isso significa que a assinatura deixada por essas duas perspectivas é perpetuada ao longo das semanas, o que facilita tanto a reidentificação desses usuários quanto a proteção, caso eles mantenham padrões semelhantes em termos de nível de vulnerabilidade.

**Tabela 3. Mudanças de perfil-vulnerabilidade ao longo das semanas**

	Mantém o mesmo estado de vulnerabilidade?											
	Uniqueness						Hiper cubo					
	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim	Não	Sim
Mantém mesmo perfil?	1-2		1-3		2-3		1-2		1-3		2-3	
Não	4,95%	52,45%	5,62%	52,58%	5,03%	50,08%	2,63%	54,77%	3,12%	55,08%	2,81%	52,29%
Sim	3,41%	39,19%	3,67%	38,12%	3,72%	41,17%	2,01%	40,60%	1,85%	39,95%	1,54%	43,36%

## 6. Conclusões

Neste trabalho, estendemos a avaliação de técnicas de quantificação de vulnerabilidade em dados de mobilidade urbana ao incorporar uma análise temporal da relação entre perfis de mobilidade e vulnerabilidade dos usuários. Utilizamos uma metodologia composta por duas etapas: (i) a identificação de perfis de mobilidade e (ii) a quantificação da vulnerabilidade, considerando duas técnicas que capturam visões complementares da vulnerabilidade individual de usuários, *uniqueness*, que captura a vulnerabilidade espaço-temporal, e o hiper cubo, que captura a vulnerabilidade comportamental dos usuários.

A principal contribuição deste estudo é uma avaliação longitudinal dessa relação perfil–vulnerabilidade, permitindo analisar como ela evolui ao longo do tempo. Para isso, avaliamos dados de telefonia móvel da cidade do Rio de Janeiro ao longo de três semanas consecutivas. Nossos resultados indicam que, embora os perfis de mobilidade dos usuários mudem com frequência ao longo das semanas, o nível de vulnerabilidade permanece relativamente estável: usuários inicialmente protegidos tendem a se manter protegidos, enquanto usuários vulneráveis tendem a continuar vulneráveis.

Além disso, ao analisar a vulnerabilidade comportamental, a qual apresenta alto grau de interpretabilidade, observamos que os usuários vulneráveis se distanciam dos seus vizinhos recorrentemente nas mesmas cinco dimensões. Esse resultado sugere que a proteção direcionada a essas dimensões pode ser suficiente para mitigar a vulnerabilidade da maioria dos usuários, oferecendo subsídios práticos para o desenvolvimento de mecanismos de proteção e anonimização mais eficientes.

Por fim, como trabalhos futuros, pretendemos expandir essa avaliação para conjuntos de dados com diferentes características, bem como aumentar o número de usuários e o período de análise, permitindo avaliar a generalização dos padrões observados.

## 7. Agradecimentos

Este trabalho faz parte do projeto Mob Sci-Dat Factory (ANR-23-PEMO-0004) no âmbito do programa France 2030. Ele foi apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pelo projeto CAPES-STIC-AMSUD 22-STIC-07 LINT.

## Referências

- Amichi, L., Viana, A. C., Crovella, M., and Loureiro, A. A. (2020). Understanding individuals' proclivity for novelty seeking. In *SIGSPATIAL*.
- Analytica, A. (2025). Location-based services market to reach over us\$ 462.65 billion in revenue by 2032. <https://shorturl.at/bhBBg>. Accessed: 2026-01-09.
- Chaves, J. C., da Silva, M. A., de Souza Alencar, R., Evsukoff, A. G., and da Fonseca Vieira, V. (2023). Human mobility and socioeconomic datasets of the rio de janeiro metropolitan area. *Data in brief*.
- De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3(1):1–5.
- do Couto Teixeira, D., Almeida, J. M., and Viana, A. C. (2021). On estimating the predictability of human mobility: the role of routine. *EPJ Data Science*, 10(1).
- Esper, J. P., Viana, A. C., and Almeida, J. M. (2024). Beauty or beast: Human behavioral insights and learning power of federated mobility prediction. In *Proc. ACM SIGSPATIAL*.
- Farzanehfar, A., Houssiau, F., and de Montjoye, Y.-A. (2021). The risk of re-identification remains high even in country-scale location datasets. *Patterns*.
- Félix, L. G., Achir, N., Kouam, A. J., Viana, A. C., and Almeida, J. M. (2025). Estimando a vulnerabilidade à exposição de usuários em dados de mobilidade. In *Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*.
- Frissen, V. (2018). Gender is calling: Some reflections on past, present and future uses of the telephone. In *The gender-technology relation*.
- Gramaglia, M., Fiore, M., Furno, A., and Stanica, R. (2021). Glove: Towards privacy-preserving publishing of record-level-truthful mobile phone trajectories. *ACM/IMS Transactions on Data Science (TDS)*.
- Green, N. (2002). On the move: Technology, mobility, and the mediation of social time and space. *The information society*, 18(4):281–292.
- Jiang, K., Shao, D., Bressan, S., Kister, T., and Tan, K.-L. (2013). Publishing trajectories with differential privacy guarantees. In *25th Int. Conf. on scientific and statistical database management*.
- Ling, R. and Yttri, B. (1999). Nobody sits at home and waits for the telephone to ring: Micro and hyper-coordination through the use of the mobile telephone.
- Liu, T., Yang, Z., Zhao, Y., Wu, C., Zhou, Z., and Liu, Y. (2018). Temporal understanding of human mobility: A multi-time scale analysis. *PloS one*, 13(11):e0207697.
- May Petry, L., Leite Da Silva, C., Esuli, A., Renso, C., and Bogorny, V. (2020). Marc: a robust method for multiple-aspect trajectory classification via space, time, and semantic embeddings. *International Journal of Geographical Information Science*.
- Pyrgelis, A., Troncoso, C., and De Cristofaro, E. (2018). Knock knock, who's there? membership inference on aggregate location data. In *NDSS*.
- Su, R., Dodge, S., and Goulias, K. G. (2022). Understanding the impact of temporal scale on human movement analytics. *Journal of Geographical Systems*, 24(3):353–388.
- Zang, H. and Bolot, J. (2011). Anonymization of location data does not work: A large-scale measurement study. In *Proc. 17th Annual International Conference on Mobile Computing and Networking*.