

Análise do impacto da pandemia de COVID-19 na mobilidade no Brasil sob uma visão semântica

Germano B. dos Santos¹, Fabrício A. Silva¹, Thais R. M. Braga Silva¹

¹Laboratório de Inteligência em Sistemas Pervasivos e Distribuídos (NESPED-LAB)
Universidade Federal de Viçosa, Florestal, MG, Brasil

germano.santos@ufv.br, fabricio.asilva@ufv.br, thais.braga@ufv.br

Resumo. Em 2020, o vírus SARS-CoV-2 se espalhou rapidamente no Brasil, implicando no distanciamento social visando a contenção das infecções. A análise da mobilidade, nessa época, foi importante para avaliar a contaminação da população brasileira. No entanto, os efeitos causados no padrão de mobilidade pelas restrições impostas durante a pandemia é ainda pouco discutido. Neste estudo, são analisados 95.522.812 registros de 4.279.025 usuários móveis, referentes aos anos de 2021 e 2022, visando compreender como a população brasileira se adaptou ao novo ambiente pós-pandemia em uma visão semântica da mobilidade humana. A partir da matriz de comparação dos padrões de mobilidade, da avaliação de clusters de usuários móveis baseados na representação vetorial de motivos semânticos e de métricas de deslocamentos, observa-se uma gradual mudança do teletrabalho para o modelo híbrido, aumentando a imprevisibilidade dos deslocamentos dos brasileiros entre 2021 e 2022.

Abstract. In 2020, the SARS-CoV-2 spread quickly in Brazil, implicating on non-pharmaceutical interventions such as social distancing. In this context, the mobility analysis was important to evaluate the infection rates of Brazilian people. Nevertheless, the possible effects on mobility patterns due to the pandemic restrictions during this period are often overlooked. In this context, it is analyzed 95,522,812 mobility records of 65,402 mobile users, regarding the years 2021 and 2022 aiming at understanding how the Brazilian people adapted to the restrictions easing under a semantic view of human mobility. Therefore, using the comparison semantic motif matrix, the analysis of users' clusters based on embeddings of semantic motifs, and mobility metrics, it is observed that there is a gradual change from teleworking to hybrid models, increasing the displacements' variability of Brazilian people from 2021 to 2022.

1. Introdução

A mobilidade da população mundial sofreu alterações após as intervenções governamentais de contingência em razão do espalhamento do vírus SARS-CoV-2 [Chinazzi et al. 2020]. O estudo dos deslocamentos durante a pandemia de COVID-19 foi importante para entender e prever o comportamento da contaminação da população mundial [Nouvellet et al. 2021]. Segundo [Song et al. 2010], as pessoas não se movimentam aleatoriamente, por possuírem preferência por locais seguindo diferentes características. Portanto, estimar como e por que se toma a decisão de qual local visitar em um

determinado horário e dia da semana é importante para a tarefa de controle de epidemias [Barbosa et al. 2018].

No entanto, o estudo da mobilidade da população se restringiu a conjuntos de dados sem detalhes semânticos. [Rowe et al. 2023] utilizam dados gerais de viagens (origem-destino) disponibilizados pelo Facebook-Meta para analisar a mobilidade na Grã-Bretanha. [Iio et al. 2021] analisam apenas métricas de deslocamento extraídas de dados móveis disponibilizados por uma empresa, nos Estados Unidos. Contudo, o padrão de mobilidade de um usuário representa seu estilo de vida, que está diretamente relacionado com o seu deslocamento diário [Ma et al. 2023]. Além disso, ainda existem poucos trabalhos que discutem os impactos causados na mobilidade humana durante e após restrições impostas considerando o Brasil como área de estudo [Chagas et al. 2021].

Portanto, este trabalho preenche essa lacuna e analisa o impacto das restrições da pandemia de COVID-19 nos anos de 2021 e 2022, através do estudo do padrão de mobilidade semântico de usuários móveis utilizando dados reais de localizações coletados a partir de GPS de 4.279.025 usuários. Esse padrão é definido como *motifs* semânticos, um subgrafo em que seus vértices representam as regiões de parada de cada usuário móvel categorizados como *Casa*, *Trabalho* ou *Outro* e suas arestas as visitas entre as regiões. Além disso, é proposto um novo método de análise coletiva de mobilidade utilizando o agrupamento de características numéricas extraídas de uma sequência de *motifs* semânticos com a aplicação do algoritmo Doc2Vec [Le and Mikolov 2014]. Com a análise semântica dos padrões de deslocamentos dos usuários móveis, notou-se que existe uma transição gradual entre o modelo de trabalho à distância para o modelo híbrido, e que os usuários voltaram a se deslocar para locais com uma semântica variada, diferente de *Casa* e *Trabalho*, aumentando a imprevisibilidade da sua rotina.

O restante deste texto está organizado da seguinte forma. Na Seção 2 são apresentados os trabalhos relacionados. A Seção 3 descreve as métricas e algoritmos que compõem o método proposto. A Seção 4 apresenta os conjuntos de dados e os resultados das análises. Na Seção 5 são discutidos os resultados de forma analítica. Por fim, na Seção 6 são apresentadas as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Entender como as pessoas se movimentam e por que interagem diferentemente conforme o tipo do estabelecimento que visitam, é um desafio [Song et al. 2010] e ainda atual [Yang et al. 2023]. Nesse contexto, a caracterização semântica da mobilidade é um tema importante para a compreensão das preferências de usuários móveis. O conceito de *motifs* é discutido inicialmente por [Schneider et al. 2013], porém [Ma et al. 2023] mostram que a partir da análise de *motifs* semânticos é possível inferir o estilo de vida dos usuários. Além disso, os *motifs* também podem indicar relações espaciais como observado em [Cao et al. 2019].

Nos últimos anos, estudos que visam analisar tendências e mudanças de comportamento de deslocamento surgiram, principalmente devido às restrições impostas pela pandemia de COVID-19 [Benita 2021]. A relação entre redes complexas e pontos de interesse indicam como foi a adaptabilidade da população americana diante das restrições da COVID-19 [Boz et al. 2024]. Além disso, a mobilidade pode caracterizar disparidades socioeconômicas como evidenciado em [Rowe et al. 2023, Iio et al. 2021].

Alguns estudos caracterizam, superficialmente, a semântica do padrão de visita de usuários móveis para avaliar a diferença de mobilidade em diferentes períodos da COVID-19. [Li et al. 2021] analisam o intuito de viagens a partir de dados de compartilhamento de bicicletas e pontos de interesse em Zurich durante a pandemia. Por outro lado, [Yao et al. 2022] analisam grupos de usuários gerados a partir da clusterização da similaridade de sequências de visitas para avaliar a diferença de comportamento nas primeiras restrições da pandemia na China. [Bouzaghane et al. 2024] propõem uma métrica de similaridade considerando uma sequência de viagens utilizando similaridade de cossenos, para compreender a mudança do comportamento da mobilidade humana, nos Estados Unidos, do período pós-pandêmico em comparação aos períodos anteriores.

No entanto, os trabalhos que analisam o impacto na mobilidade causado pelo distanciamento social no Brasil não utilizam a semântica das visitas [Fanticelli et al. 2022, Dias et al. 2024, Nepomuceno et al. 2022]. Portanto, o presente trabalho tem o intuito de extrair características dos *motifs* semânticos e métricas de mobilidade para analisar o efeito da mobilidade após o relaxamento das restrições da pandemia da COVID-19 no Brasil.

3. Metodologia

A metodologia proposta é composta de 4 etapas: (1) extração de regiões de paradas, (2) extração dos *motifs* semânticos, (3) comparação dos *motifs* frequentes entre dois períodos, (4) análise das métricas de deslocamento.

3.1. Extração de Regiões de Paradas

Ao longo deste estudo, será utilizado L_{u_i} para se referir ao conjunto de localizações do usuário $u_i \in U$. A localização $l_i \in L_{u_i}$, portanto, pode ser definida como $l_i = \langle id, lat, long, timestamp \rangle$ em que *id* representa o identificador único do usuário, a latitude, longitude e *timestamp* definem um evento geolocalizado.

[Montoliu et al. 2013] definem dois conceitos para identificar lugares de interesse de acordo com suas visitas: ponto de parada e região de parada. Um ponto de parada $lp_i \in LP_{u_i} \subset L_{u_i}$ é um agrupamento de localizações em que l_{i+1} está a no máximo D_{max} metros de distância de l_i e tenha uma diferença de tempo de no mínimo T_{min} e no máximo T_{max} .

Formalmente, os pontos de paradas podem ser definidos como $lp_i = \langle id, t_{inicio}, t_{fim}, lat, long \rangle$. O primeiro campo define o identificador do usuário, o t_{inicio} e o t_{fim} representam, respectivamente, o tempo de chegada e o tempo de saída daquela parada, e a localização geográfica é especificada por *lat* e *long*.

Uma região de parada é calculada a partir do agrupamento de pontos de paradas, gerando um retângulo mínimo que contém todos esses pontos, a partir das coordenadas geográficas mínimas e máximas de cada grupo. Portanto, uma região é definida como $r_i^u = \langle id, minbox, lat_{cm}, long_{cm} \rangle$, sendo *id*, o identificador do usuário; *minbox*, o retângulo mínimo que contém alguns pontos de paradas LP_{u_i} ; latitude (lat_{cm}) e longitude ($long_{cm}$) representam o centroide da região r_i^u . A Figura 1 representa a identificação de uma região de parada de um usuário móvel.

Nesse estudo, assume-se três categorias semânticas possíveis para uma região de parada: Casa, Trabalho e Outro. Para classificar uma região de parada r é necessário defi-

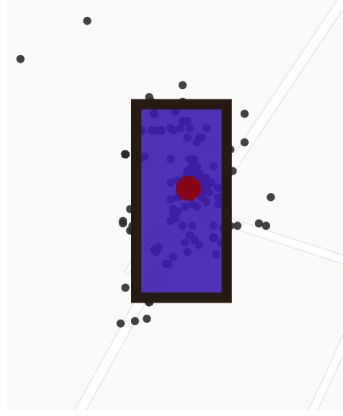


Figura 1. Exemplo da região de parada de um usuário u utilizando a definição apresentada acima. O retângulo mínimo que contém os pontos de parada LP_{u_i} é representado pela cor azul e preta. Os pontos de parada lp_i^u que não estão contidos no retângulo possuem $d > D_{max}$ ou $t < T_{min}$ ou $t > T_{max}$. O ponto vermelho representa o centroide desse polígono.

nir os horários de visita mais frequentes. Segundo [Capanema et al. 2019], cada usuário possui uma rotina diferente e por isso a classificação semântica das regiões de parada deve ser personalizada, ou seja, relacionada com a rotina do usuário móvel. Portanto, é preciso considerar os padrões de visita singulares em cada região de visita para a classificação personalizada. Dessa forma, a região r_i^u é definida como Trabalho se a maioria das visitas do usuário u , que acontecem na região r_i^u , pertencem ao intervalo de 9:00 às 19:00; a definição de casa segue a mesma definição, mas para o intervalo de 20:00 às 8:00. Se a região não atender às restrições anteriores, será classificada como "Outro".

3.1.1. Extração dos *Motifs* Semânticos

Para extrair o comportamento semântico de um usuário móvel, é necessário analisar os padrões de visita, caracterizados por *motifs* semânticos. Um *motif* é um subgrafo direcionado recorrente que representa as visitas e deslocamentos em um intervalo de tempo [Schneider et al. 2013]. O *motif* semântico é rotulado a partir da semântica do local visitado (neste estudo, em Casa, Trabalho ou Outro). Para simplificar, foram usadas as abreviações H para Casa, W para Trabalho e O para Outro. Então, um usuário com *motif* semântico 'HWH' tende a sair de casa (H), ir para o trabalho (W) e voltar para casa (H), como mostrado no primeiro exemplo da Figura 2. [Xiong et al. 2021] definem *motifs* descontínuos como trajetórias que não possuem ciclo, ou seja, não retornam ao seu local de início, como exemplificado nos dois últimos *motifs* da Figura 2.

Neste trabalho, os *motifs* de cada usuário são identificados conforme os seus respectivos pontos de parada. Para tanto, cada ponto de parada lp_i^u pode ser rotulado conforme sua relação espacial com uma região de parada r_i^u . Se o ponto lp_i^u está contido na região r_i^u , o ponto de parada recebe o mesmo rótulo dessa região, caracterizando uma visita. Dessa forma, um *motif* semântico pode ser escrito como $\mathcal{M}(V, A)$, onde V é o conjunto de regiões visitadas $v_i = \langle lat, long, cat \rangle$, representados por latitude, longitude e uma categoria, e A é o conjunto de arestas ordenado temporalmente, que representa os deslocamentos de um local v_i a outro v_{i+1} . Com isso, é possível agregar esses *motifs*

para que se tenha o padrão mais frequente para cada usuário, e o mais frequente para uma região.

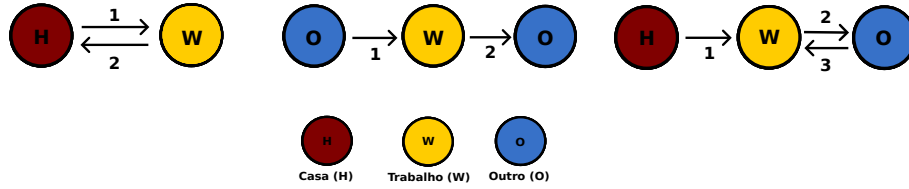


Figura 2. Exemplos de *motifs* semânticos. O primeiro *motif* da esquerda para direita representa os deslocamentos Casa-Trabalho-Casa ('HWH'). O segundo *motif* caracteriza o deslocamento Outro-Trabalho-Outro ('OWO'), nesse caso, o usuário pode ter ido à Farmácia, depois ao Trabalho e Restaurante. O terceiro *motif* representa 3 deslocamentos com quatro localizações, Casa-Trabalho-Outro-Trabalho ('HWOW').

3.1.2. Comparação de *motifs* frequentes entre períodos

Esta seção apresenta a formalização da matriz que relaciona a proporção entre os *motifs* mais frequentes considerando dois períodos distintos. A matriz pode representar a mobilidade em uma visão microscópica. Além disso, essa matriz pode representar a mobilidade macroscópica, em uma visão regional.

Considerando a visão microscópica, a função binária $\theta_{m(i,d1),m(j,d2)}(u) : M_{d1} \times M_{d2} \rightarrow \{0, 1\}$ define se o usuário u apresenta a transição do *motif* $m(i,d1) \in M_{d1}$ para o *motif* $m(j,d2) \in M_{d2}$, em que M_{d1} e M_{d2} são os conjuntos que representam os *motifs* mais frequentes de cada conjunto de dados $d1$ e $d2$. Portanto, a proporção da transição, de usuários que mudaram do *motif* mais frequente $m(i,d1)$ no conjunto de dados $d1$ para o *motif* $m(j,d2)$, no conjunto de dados $d2$, é calculada por $\alpha(m(i,d1), m(j,d2)) = \frac{|U_{m(i,d1)}|}{\sum_{m(i,d1)} \theta_{m(i,d1),m(j,d2)}(u_i) / |U_{m(i,d1)}|}$, em que $U_{m(i,d1)}$ é o conjunto dos usuários que possuem o *motif* $m(i,d1)$ como o mais frequente no conjunto $d1$.

Assim, para cada *motif* pertencente a M_{d1} , temos um vetor $\mathcal{V}_{m(i,d1)} = [\alpha(m(i,d1), m(j,d2)), \alpha(m(i,d1), m(j+1,d2)), \dots, \alpha(m(i,d1), m(n,d2))]$ que define a proporção da mudança de um *motif* $m_i \in M_{d1}$ para todos *motifs* contidos em M_{d2} . Então, é possível definir a matriz de transição $\mathcal{M}_{(d1,d2)} = [\mathcal{V}_{m(i,d1)}^T, \mathcal{V}_{m(i+1,d1)}^T, \dots, \mathcal{V}_{m(n,d1)}^T]$.

Similarmente, é possível analisar a visão macroscópica de transição por região. Para isso, cada usuário é mapeado à região que ele mais visita considerando um conjunto de dados $\mathcal{R}(u)_{d1} = \arg \max_r P_u : U \times R \rightarrow R$, em que P_u é uma função que calcula a quantidade de pontos de parada de cada usuário u , no conjunto $d1$, na região r , sendo que R é o conjunto de todas as regiões que compõem a área de estudo. Note que essa função também deve ser aplicada para o conjunto $d2$, para existir a transição. Assim, temos os conjuntos M_{d1}^r, M_{d2}^r dos *motifs* mais frequentes em cada região r em cada período. Utilizando a mesma ideia da função binária anterior, temos $\gamma_{m(i,d1),m(j,d2)}(r) : M_{d1}^r \times M_{d2}^r \rightarrow \{0, 1\}$ que define se a região r apresenta a transição do *motif* $m(i,d1) \in M_{d1}^r$ para o *motif* $m(j,d2) \in M_{d2}^r$. Logo, $\delta(m(i,d1), m(j,d2)) = \frac{|R_{m(i,d1)}|}{\sum_{m(i,d1)} \gamma_{m(i,d1),m(j,d2)}(r_i) / |R_{m(i,d1)}|}$. A

definição da matriz de transição para região é semelhante ao do usuário, apenas alterando a função que calcula a proporção da transição. Note ainda que para a agregação dos *motifs* por região não é necessário ter os mesmos usuários nas duas bases de dados que serão analisadas.

3.1.3. Métricas de Deslocamento

Além dos *motifs* semânticos, neste trabalho o estudo do efeito da mobilidade engloba as métricas de: 1) entropia não-correlacionada ao tempo [Pappalardo et al. 2016], uma métrica de mobilidade baseada na função de probabilidade de um usuário visitar uma localização dentre todos os distintos locais previamente conhecidos; 2) raio de giro [Gonzalez et al. 2008], que indica a distância característica de um usuário; e 3) k -raio de giro [Pappalardo et al. 2015], que indica o raio de giro a partir dos k locais mais visitados por um usuário.

A partir da definição de [Pappalardo et al. 2015] é possível classificar os usuários em *returners* ou *explorers*, baseado na relação entre raio de giro e o k -raio de giro. Se o k -raio de giro é menor ou igual ao raio de giro dividido por dois, então o usuário é classificado como *returner*, porque é possível explicar o deslocamento característico desse usuário a partir das suas k localizações mais visitadas; senão é classificado como *explorer*. Um usuário *returner* é considerado conservador, que visita locais já conhecidos, enquanto o *explorer* visita locais novos frequentemente.

4. Resultados

Nesta seção, o impacto das restrições de mobilidade governamentais devido à COVID-19 é analisado considerando uma visão semântica dos deslocamentos, a partir de milhões de dados coletados de milhões de brasileiros. O principal objetivo desta análise é identificar e discutir os impactos semânticos nos deslocamentos diários.

4.1. Dados e Execução

Foram utilizados dados reais coletados a partir de GPS, disponibilizados por uma empresa parceira sob acordo de confidencialidade. A coleta foi realizada de forma não intermitente, portanto, a trajetória de alguns usuários podem ser descontínuas. A base utilizada neste trabalho é constituída de 65.402 usuários, coletados em 2021, de 1 de abril a 30 de maio; os dados de 2022 possuem registros de 4.213.623 usuários, de 1 de outubro a 30 de novembro. Os dados de 2021 são compostos por 49.435.713 de registros e os de 2022 contêm 46.087.099 de registros, totalizando 95.522.812 de registros. Vale ressaltar que os métodos propostos e as análises feitas neste trabalho focam em padrões coletivos de mobilidade, e, portanto as bases de 2021 e 2022 não precisam contemplar necessariamente os mesmos usuários.

Para seguir a Lei Geral da Proteção de Dados (LGPD), os usuários foram anonimizados e os identificadores criptografados.

A comparação do comportamento dos usuários será discutida nas próximas seções, porém deve-se considerar a esparsidade dos registros de 2022. Nota-se uma diferença na quantidade de registros média diária dos usuários de 2022 em relação a 2021.

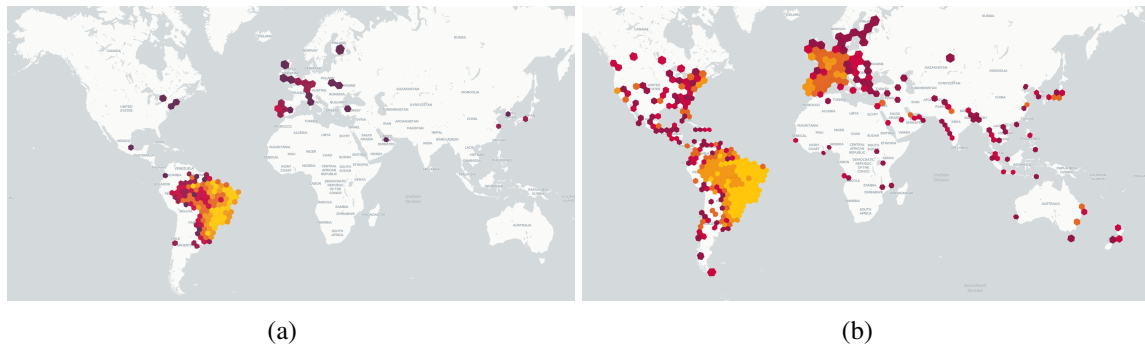


Figura 3. (a) Distribuição espacial dos pontos de paradas em 2021; (b) Distribuição espacial dos pontos de paradas em 2022;

A mediana da quantidade diária média de registros em 2021 é de, aproximadamente, 16 registros, enquanto em 2022 é de apenas 2 registros.

A Figura 3 mostra a diferença da distribuição espacial entre os anos de 2021 e 2022. No último ano nota-se uma maior dispersão, com localizações presentes em diferentes continentes. Em 2021, existe uma concentração de localizações no Brasil, devido às restrições de fluxos migratórios impostas pelo governo durante a pandemia.

Para identificar os pontos de parada dos usuários conforme a Seção 3.1, os parâmetros D_{max} , T_{min} e T_{max} foram definidos como 200 metros, 10 minutos e 8 horas, respectivamente. O parâmetro T_{min} foi definido como 10 minutos para retirar paradas em uma trajetória que não são importantes para a comparação da assinatura de mobilidade, por exemplo, como uma parada em semáforo; o parâmetro T_{max} foi definido como 8 horas, pelo fato de as pessoas, em geral, permanecerem no trabalho por aproximadamente 8 horas. Por fim, o parâmetro D_{max} é igual a 200 metros pelo fato de ser o menor valor considerado por [Montoliu et al. 2013] que produz os melhores resultados.

Para realizar o agrupamento dos pontos de parada em regiões de parada, conforme Seção 3.1, foi utilizado o algoritmo *DBSCAN*, que desconsidera ruídos, com parâmetro raio igual a $100m$. Os parâmetros que especificam o mínimo de pontos em cada agrupamento e *eps* que define a distância mínima entre os pontos vizinhos seguem os valores apresentados por [Capanema et al. 2019], 18% da quantidade de pontos de parada e 20 metros, respectivamente.

Foi utilizada uma máquina com um processador Intel i7-1165G7 (12M Cache, 2.8 GHz, 8 cores, 2 threads), 16 GB DDR4 3200 MHz RAM 512 GB de HD. A complexidade dessa extração é regida pela quantidade de usuários, devido ao algoritmo de clusterização utilizado para definir as regiões.

4.2. Métricas de Deslocamento

Primeiramente, serão analisadas as métricas de raio de giro, entropia não-correlacionada ao tempo e k -raio de giro. A Figura 4(a) apresenta a distribuição do raio de giro médio diário para os dois períodos (2021 e 2022). É possível observar que, apesar de a mediana dos valores estarem próximos, em 2022 a variação é significativamente maior. Uma hipótese para essa diferença é que em 2022 os usuários passaram a visitar mais locais, mesmo que sejam próximos do seu centro de massa. A quantidade maior de usuários no

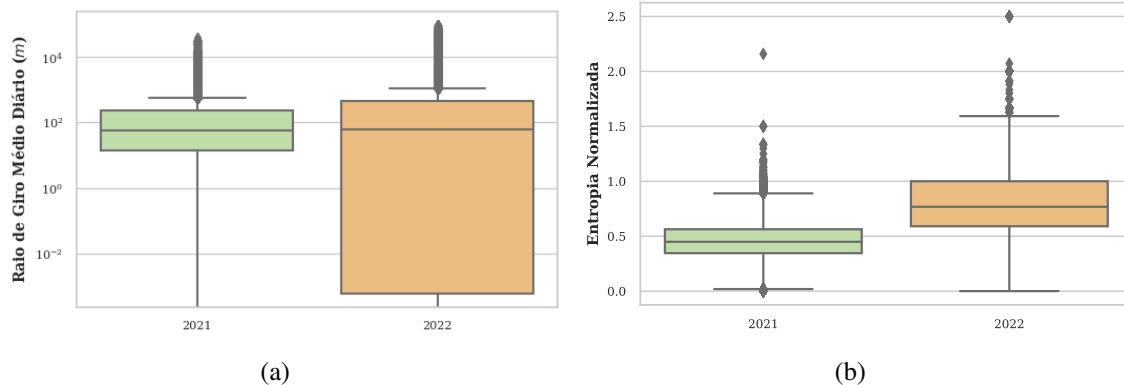


Figura 4. (a) Distribuição do raio de giro médio; (b) Distribuição da entropia não correlacionada ao tempo;

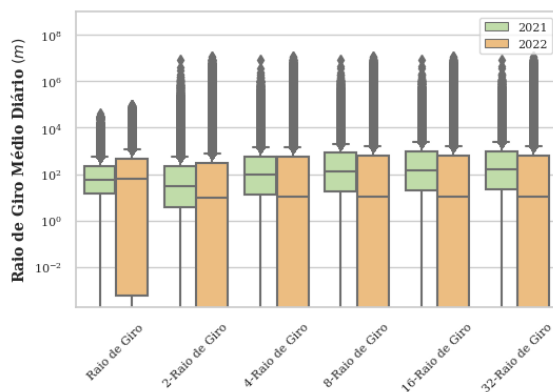


Figura 5. Distribuição do k -raio de giro nos anos de 2021 e 2022.

Ano	k	<i>Returners</i>	<i>Explorers</i>
2021	2	47,79%	52,21%
2021	4	67,78%	32,22%
2021	8	73,30%	26,70%
2021	16	74,24%	25,76%
2021	32	74,31%	25,69%
2022	2	32,59%	67,41%
2022	4	40,91%	59,09%
2022	8	41,99%	58,01%
2022	16	42,05%	57,95%
2022	32	42,06%	57,94%

Tabela 1. Quantidade de usuários classificados em *Returners* e *Explorers*

conjunto de dados de 2022 também é um fator que afeta essa métrica.

A Figura 4(b) apresenta a entropia dos usuários. A entropia dos usuários no ano de 2022 é maior que a de 2021, o que indica uma menor previsibilidade do deslocamento. Portanto, os padrões de visitas em 2022 são mais difíceis de serem previstos, pois os usuários se deslocam com maior variabilidade ao longo dos dias. Enquanto em 2021 a maioria das pessoas ainda estava seguindo alguma medida de restrição devido à pandemia. Em contrapartida, em 2022 essas pessoas passaram a se deslocar mais, mesmo que seja para locais próximos.

Além disso, realizou-se o cálculo do k -raio de giro com k sendo igual a 2, 4, 8, 16 e 32. Na Figura 5, é notável que o k -raio de giro tende a ser menor em 2022. Como mais locais são visitados, os k locais mais frequentes tendem a ter uma relevância menor. Devido ao maior número de usuários no conjunto de dados de 2022, a variabilidade também é maior nesse ano.

Analisando a Tabela 1, é possível observar que, em 2021, a relação de usuários classificados como *returners* é maior do que *explorers*, o que mostra que os usuários

tendem a retornar a locais já conhecidos. Por outro lado, essa relação se inverte em 2022, sendo a proporção de *explorers* maior, indicando que os usuários passaram a frequentar mais locais após a flexibilização das restrições da pandemia.

4.3. *Motifs* Semânticos

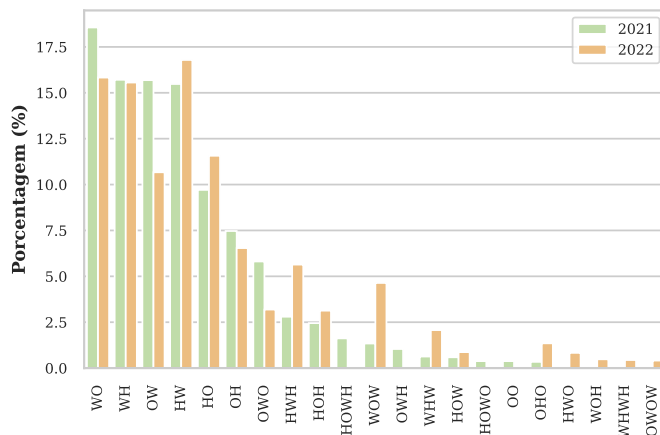


Figura 6. Frequência dos *motifs* semânticos

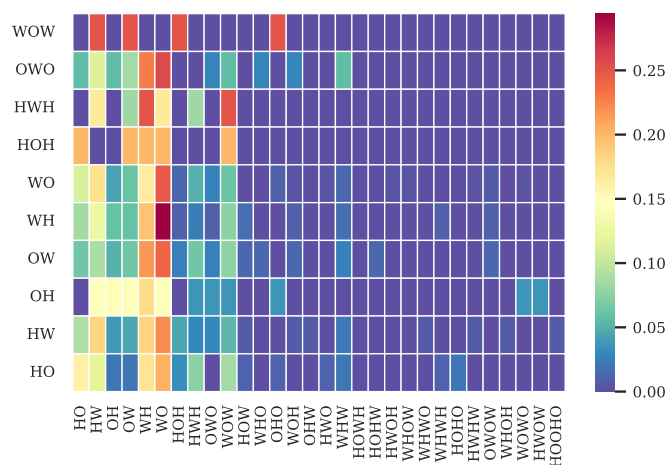


Figura 7. Transição dos *motifs* semânticos mais frequentes em 2021 e 2022. Eixo *y* representa os *motifs* frequentes de 2021, e o eixo *x* os de 2022. Os valores apresentados indicam a proporção das regiões que tinham um padrão como mais frequente em 2021 e passaram a ter outro como mais frequente em 2022.

Ao analisar a Figura 6 com a frequência de cada padrão de *motifs* semânticos, notam-se algumas mudanças. Em 2021 é possível observar a predominância de deslocamentos característicos de Casa-Trabalho ('HW') e Trabalho-Casa ('WH'). Em contrapartida, é possível notar a diversificação dos *motifs* em 2022, com padrões que não aparecem como mais frequentes em 2021, como 'OHO', 'HWO', 'WOH'. É perceptível também a influência do ano no padrão de deslocamento 'WOW', sendo em 2022 um padrão mais frequente em relação a 2021, esse padrão que pode ser exemplificado como um deslocamento no horário de almoço, o que não era comum durante a pandemia de COVID-19.

Com essas diferenças entre os padrões, é interessante conhecer os *motifs* semânticos mais frequentes de cada região e observar se mudaram ao longo desse intervalo de tempo, após as restrições impostas pela COVID-19. A Figura 7 mostra a transição dos *motifs* mais frequentes das regiões do ano 2021 para 2022. Observa-se que a principal diferença do ano de 2021 para o ano de 2022 foi mudança do padrão ‘WH’, para ‘WO’. Nesse caso, essa alteração relata a mudança de comportamento dos usuários em 2022 que preferem sair após o horário de trabalho ao invés de voltar para suas casas. A alteração do padrão ‘HWH’ para ‘HWO’ também segue a mesma característica.

Além desse comportamento importante ressaltado nessa matriz, é observável que aconteceu a diversificação de padrões de deslocamento, como também mostrado na Figura 6. Nesse caso, essas alterações podem ser identificadas na parte direita da matriz de transição. *Motifs* como, ‘OHO’, ‘HWOW’, ‘HOHO’, são em 2022 padrões que se tornaram mais frequentes da rotina de alguns usuários. Note que em 2021, nenhum desses padrões são observados como mais frequentes. Além disso, essa diversificação também pode ser vista pela quantidade de deslocamentos diários, visto que em 2022 alguns *motifs* mais frequentes são maiores em tamanho quando comparado a 2021.

4.3.1. Agrupamento dos *Motifs*

A análise dos grupos dos padrões de mobilidade de cada período tem o objetivo de investigar as características dos deslocamentos em cada ano. Primeiramente, é necessário extrair características numéricas dos *motifs* semânticos de cada usuário. [Yao et al. 2022, Bouzaghane et al. 2024] consideram a similaridade entre a sequência de deslocamento, contudo não capturam as características latentes em um padrão de mobilidade complexo de um usuário móvel [Song et al. 2010].

Portanto, para esta análise, adotou-se que cada *motif* é uma ‘palavra’ e a sequência de *motifs* é um ‘texto’. Sendo assim, é possível extrair a representação vetorial do padrão de mobilidade do usuário com o algoritmo *Doc2Vec* [Le and Mikolov 2014]. Nesse sentido, um usuário u é caracterizado pelo padrão de mobilidade extraído dos *motifs* semânticos.

Após a aplicação do algoritmo *Doc2Vec*, a sequência ordenada temporalmente, $\mathbf{vm}^u = \{m_i^u, m_{i+1}^u, \dots, m_n^u\}$, em que m_i^u é o i -ésimo *motif* do usuário u , é representada pelo vetor característico $\mathbf{vm}_{\text{emb}}^u \in \mathbb{R}^d$, onde d é a dimensão do vetor $\mathbf{vm}_{\text{emb}}^u$.

Dessa forma, para esta análise, os anos e seus respectivos usuários foram agrupados de acordo com cada vetor característico de *motifs*, utilizando o algoritmo *K-Means* com k igual a 2. O k foi escolhido com base na silhueta do agrupamento, que determina a proporção da distância entre os elementos do mesmo grupo e os elementos de grupos distintos. Para $k = 2$, a silhueta é igual a 0,32.

Para o ano de 2021, o grupo 0 possui 71,54% dos usuários, enquanto o grupo 1 possui 28,46% dos usuários. O grupo 0 em 2021 é caracterizado por usuários que possuem deslocamentos mais conservadores. Analisando a composição dos *motifs* desse grupo, nota-se que 93% possuem deslocamentos para o Trabalho, 66% para Casa e 43% para outra localização. Para o grupo 1, 63,3% dos *motifs* possuem deslocamentos para o Trabalho, 51,2% para Casa e 92,6% para outra localização. Ainda é possível observar que

o tamanho de um *motif* também é importante para diferenciar os grupos; para o grupo 0, 90,7% representam apenas 1 deslocamento (e.g., HW) e 6% representam 2 deslocamentos (e.g., OWO). Para o grupo 1, 68,7% dos *motifs* possuem 1 deslocamento e 26,6% representam 2 deslocamentos.

Em 2022, a configuração dos grupos se inverte. O grupo 0 possui 36,14% dos usuários, enquanto o grupo 1 possui 63,86% dos usuários. A partir da análise da composição dos *motifs*, observa-se a semelhança entre os grupos do ano de 2021. Para o grupo 0, 97,3% dos *motifs* possuem deslocamentos para o Trabalho, 79,5% apresenta deslocamentos para Casa e 28,7% para outra localização. Para o grupo 1, 67,1% dos *motifs* possuem deslocamentos para o Trabalho, 59,5% apresenta deslocamentos para Casa e 77,6% para outra localização. Com isso, nota-se uma diversificação dos destinos na mobilidade dos usuários do grupo 1. Esse comportamento também é observado no tamanho dos *motifs*. Para o grupo 0, 87,2% dos *motifs* possuem 1 deslocamento e 7,4% possuem 2; enquanto para o grupo 1, 65,8% dos *motifs* possuem 1 deslocamento e 29,6% possuem 2 deslocamentos.

5. Discussão

Com os resultados obtidos da análise semântica da mobilidade dos usuários, é possível considerar os motivos da mudança do padrão de deslocamento após as restrições da pandemia de COVID-19 serem amenizadas.

O raio de giro de 2022 é parecido com o raio de giro de 2021, porém existiram muitas mudanças semânticas em 2022. Esse resultado pode ser explicado pelo fato de usuários terem mudado o seu padrão de deslocamento após a pandemia de COVID-19. Durante a pandemia era necessário conhecer lugares próximos a sua casa/trabalho, para que o vírus não se espalhasse, observação também feita em [Iio et al. 2021].

Analisando a Figura 4(b), observa-se que a entropia dos usuários, em geral, aumentou em 2022, o que demonstra que a previsibilidade de suas rotinas diminuiu. Esse comportamento pode ser corroborado pelos percentuais da Tabela 1 que mostra que existe uma maior parcela de usuários exploradores em 2022 do que em 2021 para todos os valores de k . Tal fato representa que usuários mudaram o seu comportamento, visitando muitos locais desconhecidos pela primeira vez, ou seja, se deslocando para locais que não são Casa e Trabalho.

Os *motifs* semânticos frequentes de 2021 e de 2022 possuem diferenças. O padrão ‘HW’ em 2021 é o quarto mais frequente, já em 2022 é o primeiro. Essa alteração pode ser entendida também como uma variação nos padrões de mobilidade do usuário. Quando se analisa essa mudança juntamente à matriz de transição, observa-se uma alteração nos padrões de trabalho para a gradual mudança de *home office* e educação à distância para um modelo híbrido, como também observado em [Santana et al. 2023, Rowe et al. 2023]. Durante a pandemia de COVID-19, foi necessário a adoção de trabalho remoto ou modelo de trabalho híbrido.

A mudança do padrão de mobilidade com o relaxamento das restrições da pandemia do COVID-19 é mostrado também pela análise dos grupos dos *motifs* semânticos dos períodos analisados. Nota-se que em 2021 os usuários são mais conservadores, o grupo 0, constituído por *motifs* caracterizados com mais deslocamentos Casa-Trabalho,

possui mais usuários que o grupo explorador. O comportamento explorador, grupo 1, no ano de 2021, pode ser explicado pela condição socio-econômica relatada por [Lio et al. 2021, Boz et al. 2024], pessoas que possuíam uma alta renda per capita conseguiram se adaptar às imposições da pandemia e reduzir a mobilidade. Em 2022, com o relaxamento das restrições, é possível perceber uma diversificação da mobilidade entre os usuários, o grupo explorador se torna o grupo com maior quantidade de usuários móveis. Além disso, no grupo conservador uma maior parte dos *motifs* possuem deslocamentos para uma localização diferente de Casa e Trabalho. Portanto, nesse período, os usuários começaram a se deslocar mais e para diferentes lugares.

Portanto, a análise do padrão de deslocamento dos usuários móveis revela que a mobilidade mudou para um comportamento menos previsível devido ao relaxamento das restrições impostas pelo governo durante a pandemia. O impacto da pandemia na mobilidade, portanto, é entendido como não permanente devido à população mudar seu comportamento conservador visto em 2021 para um comportamento explorador observado em 2022. Esses resultados se mostram consoantes ao estudo feito em Barcelona, Espanha [Nello-Deakin et al. 2024], demonstrando que a análise semântica do deslocamento é importante para a extração de conhecimento de padrões de mobilidade.

6. Conclusão

Nesse trabalho, foi proposto um método de extração de *motifs* semânticos a partir de dados móveis coletados provenientes de GPS. Além disso, foi definido, formalmente, uma região de parada e sua classificação semântica. Com esse método é possível analisar o padrão de mobilidade em um período, ou utilizar a definição da matriz de transição, para avaliar a mudança do comportamento do usuário em dois períodos distintos. Ainda foi apresentado como extrair características numéricas a partir da sequência de *motifs* semânticos, para analisar o comportamento dos usuários móveis em um nível coletivo. Portanto, diante dos resultados apresentados, nota-se a importância do estudo da mobilidade dos usuários para entender o contexto de sua rotina e preferências ao se deslocar.

No entanto, a semântica é ainda um grande desafio. Apesar dessa análise ser fundamental para o entendimento do comportamento da mobilidade do usuário, não é claro definir como o usuário alterou sua rotina. É preciso ainda identificar mais categorias de locais, ou seja, explorar mais profundamente a semântica. Com essa definição, é possível traçar um perfil mais assertivo dos padrões de comportamento móvel das pessoas. Portanto, como trabalhos futuros, pretende-se propor um novo método que classifique o tipo do local visitado em diversas categorias. Assim, será possível ter maior assertividade em determinar a preferência do usuário por um tipo de local específico, como comércio, alimentação e educação.

Agradecimento

Os autores gostariam de agradecer, pelo suporte financeiro, as instituições Capes, CNPq e FAPEMIG.

Referências

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.

- Benita, F. (2021). Human mobility behavior in covid-19: A systematic literature review and bibliometric analysis. *Sustainable Cities and Society*, 70:102916.
- Bouzaghane, M. A., Obeid, H., González, M., and Walker, J. (2024). Human mobility reshaped? deciphering the impacts of the covid-19 pandemic on activity patterns, spatial habits, and schedule habits. *EPJ Data Science*, 13(1):1–20.
- Boz, H. A., Bahrami, M., Balcisoy, S., Bozkaya, B., Mazar, N., Nichols, A., and Pentland, A. (2024). Investigating neighborhood adaptability using mobility networks: a case study of the covid-19 pandemic. *Humanities and Social Sciences Communications*, 11(1):1–11.
- Cao, J., Li, Q., Tu, W., and Wang, F. (2019). Characterizing preferred motif choices and distance impacts. *Plos one*, 14(4):e0215242.
- Capanema, C. G., Silva, F. A., and Silva, T. R. M. (2019). Identificação e classificação de pontos de interesse individuais com base em dados esparsos. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 15–28. SBC.
- Chagas, E. T. C., Barros, P. H., Cardoso-Pereira, I., Ponte, I. V., Ximenes, P., Figueiredo, F., Murai, F., Couto da Silva, A. P., Almeida, J. M., Loureiro, A. A. F., and Ramos, H. S. (2021). Effects of population mobility on the covid-19 spread in brazil. *PLOS ONE*, 16(12):1–27.
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastor y Piontti, A., Mu, K., Rossi, L., Sun, K., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368(6489):395–400.
- Dias, E., Diniz, A. M., Souto, G. R., Guerra, H. L., Marques-Neto, H. T., Malinowski, S., and Guimarães, S. J. F. (2024). Predicting covid-19 cases in belo horizonte—brazil taking into account mobility and vaccination issues. *Plos one*, 19(2):e0269515.
- Fanticelli, H. C., Rabenjamina, S., Viana, A. C., Stanica, R., De Oliveira, L. S., and Ziviani, A. (2022). Data-driven mobility analysis and modeling: Typical and confined life of a metropolitan population. 8(3).
- Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *nature*, 453(7196):779–782.
- Iio, K., Guo, X., Kong, X., Rees, K., and Wang, X. B. (2021). Covid-19 and social distancing: Disparities in mobility adaptation between income groups. *Transportation Research Interdisciplinary Perspectives*, 10:100333.
- Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.
- Li, A., Zhao, P., Haitao, H., Mansourian, A., and Axhausen, K. W. (2021). How did micro-mobility change in response to covid-19 pandemic? a case study based on spatial-temporal-semantic analytics. *Computers, environment and urban systems*, 90:101703.
- Ma, J., Li, B., and Mostafavi, A. (2023). Characterizing urban lifestyle signatures using motif properties in network of places. *Environment and Planning B: Urban Analytics and City Science*, page 23998083231206171.

- Montoliu, R., Blom, J., and Gatica-Perez, D. (2013). Discovering places of interest in everyday life from smartphone data. *Multimedia tools and applications*, 62(1):179–207.
- Nello-Deakin, S., Diaz, A. B., Roig-Costa, O., Miralles-Guasch, C., and Marquet, O. (2024). Moving beyond covid-19: Break or continuity in the urban mobility regime? *Transportation Research Interdisciplinary Perspectives*, 24:101060.
- Nepomuceno, T. C. C., Garcez, T. V., Silva, L., and Coutinho, A. P. (2022). Measuring the mobility impact on the covid-19 pandemic. *Mathematical Biosciences and Engineering*, 19(7):7032–7054.
- Nouvellet, P., Bhatia, S., Cori, A., Ainslie, K. E., Baguelin, M., Bhatt, S., Boonyasiri, A., Brazeau, N. F., Cattarino, L., Cooper, L. V., et al. (2021). Reduction in mobility and covid-19 transmission. *Nature communications*, 12(1):1–9.
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature communications*, 6(1):1–8.
- Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., and Giannotti, F. (2016). An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics*, 2(1):75–92.
- Rowe, F., Calafiore, A., Arribas-Bel, D., Samardzhiev, K., and Fleischmann, M. (2023). Urban exodus? understanding human mobility in britain during the covid-19 pandemic using meta-facebook data. *Population, Space and Place*, 29(1):e2637.
- Santana, C., Botta, F., Barbosa, H., Privitera, F., Menezes, R., and Di Clemente, R. (2023). Covid-19 is linked to changes in the time–space dimension of human mobility. *Nature Human Behaviour*, 7(10):1729–1739.
- Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., and González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246.
- Song, C., Koren, T., Wang, P., and Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature physics*, 6(10):818–823.
- Xiong, Q., Liu, Y., Xie, P., Wang, Y., and Liu, Y. (2021). Revealing correlation patterns of individual location activity motifs between workdays and day-offs using massive mobile phone data. *Computers, Environment and Urban Systems*, 89:101682.
- Yang, Y., Pentland, A., and Moro, E. (2023). Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics. *EPJ Data Science*, 12(1):1–15.
- Yao, W., Yu, J., Yang, Y., Chen, N., Jin, S., Hu, Y., and Bai, C. (2022). Understanding travel behavior adjustment under covid-19. *Communications in Transportation Research*, 2:100068.