

Análise da Mobilidade Urbana por meio de Redes Sociais Baseadas em Localização: Estudo de Caso em Cidades Inteligentes

João A. S. Silva, Felipe D. Cunha, Silvio Jamil F. Guimarães

¹Instituto de Ciências Exatas e Informática
Pontifícia Universidade Católica de Minas Gerais
R. Dom José Gaspar, 500 – Belo Horizonte – MG – Brazil

joao.silva.452811@sga.pucminas.br, {felipe,sjamil}@pucminas.br

Abstract. *The number of people using social media is growing every day, and as a result, the amount of data available about users is also increasing. Currently, about 4,8 billion people use social media worldwide, which is approximately 59% of the global population. Given this scenario, the literature contains works that address techniques for data collection and sampling from social media, allowing for the interpretation of such data for analysis in different domains, such as urban mobility, region classification, among others. This study utilizes Location-Based Social Networks (LBSNs) to analyze human mobility within urban centers. Check-ins collected from around the world were used to identify distinct behavior patterns among inhabitants from different regions of the planet, based on the concentration of records in specific locations. The results indicate that, despite the disparity in the spatial distribution of the data, LBSNs are capable of depicting the reality of cities, even across different cultures, making them a valuable means for collecting data for smart cities due to data availability and the easy scalability of their applications.*

Resumo. *O número de pessoas que fazem uso de redes sociais cresce a cada dia, com isso, a quantidade de dados disponíveis a cerca dos usuários também aumenta. Atualmente, cerca de 4,8 bilhões de pessoas fazem uso das redes sociais em todo o mundo, o que equivale a cerca de 59% da população mundial. Tendo em vista este cenário, na literatura podem ser encontrados trabalhos que abordam técnicas para coleta e amostragem de dados advindo das redes sociais que permitem a interpretação dos mesmos para análises de diferentes domínios, como por exemplo, mobilidade urbana, classificação de regiões, entre outros. Este trabalho faz uso de Redes Sociais baseadas em Localização (LBSN) para analisar a mobilidade humana dentro de centros urbanos. Foram coleados check-ins dispostos pelo mundo, foi possível a identificação do comportamento distinto entre habitantes de diferentes regiões do planeta, de acordo com a concentração de registros em determinadas localidades. Resultados indicam que, embora a disparidade na disposição espacial dos dados, as LBSNs são capazes de retratar a realidade das cidades mesmo em culturas diferentes, mostrando-se um bom meio para a coleta de dados voltados para cidades inteligentes, devido a disponibilidade dos dados e a fácil escalabilidade de suas aplicações.*

1. Introdução

Nos últimos anos houve um aumento considerável no número de usuários das redes sociais, onde aproximadamente 4,8 bilhões de pessoas possuem cadastro ativo nas mesmas, equivalente a cerca de 59% da população mundial. O grande número de usuários implica em um grande volume de dados gerado diariamente, dados que possuem usabilidade para diferentes domínios, tendo como foco no presente trabalho, o seu uso como sensor participativo dentro da computação urbana, voltado para a análise da mobilidade humana nas cidades, a partir da coleta de dados contendo informações espaço-temporais dos usuários.

Voltando para uma rede social mais específica, o *Foursquare*¹ possui aproximadamente 50 milhões de usuários espalhados pelo mundo, sendo reconhecido pelo trabalho com localização, possuindo funcionalidades como recomendações de acordo com a localização do usuário, guia de locais próximos ao usuário, possibilidade de registrar *check-ins*, entre outras funcionalidades. *Check-ins* são *posts* associados à localização real do usuário nas redes sociais, que possuem tal funcionalidade. O *Foursquare* possui uma plataforma dedicada especialmente aos *check-ins*, chamada *Foursquare Swarm*, que possui mais de 14 bilhões de *check-ins* registrados desde o ano de 2010, que podem ser utilizados como meio de sensoriamento participativo, assim como feito em [Silva et al. 2014] no meio da computação urbana.

Assim como dito em [Silva et al. 2023], a computação urbana faz uso de grande quantidade de fontes de dados, como dispositivos de Internet das Coisas (IoT), dados de Redes Sociais Baseadas em Geolocalização (*LBSN*) e também dados estatísticos que facilitam na compreensão do ambiente urbano. A partir da análise dos dados obtidos é possível trabalhar com dados advindos de diferentes domínios, assim como é mostrado em [Machado et al. 2015] e em [Gubert et al. 2022]. A partir dos dados obtidos, aplicações voltadas para o sensoriamento participativo dentro da área da computação urbana são capazes de realizar o monitoramento de regiões, facilita na compreensão das áreas urbanas com baixo custo, sendo capaz também de detectar a influência de fatores externos, como por exemplo, as condições climáticas e informações temporais impactando na mobilidade urbana e na concentração de pessoas em locais distintos.

Este trabalho busca por meio dos dados obtidos através do *Foursquare* mapear possíveis aplicações para os dados coletados a partir desta *LBSN*. Para isso, são explorados meios para realizar a coleta dos dados de forma a não infringir nenhuma das políticas de privacidade das redes provedoras dos dados. A partir dos dados coletados, são propostas possíveis aplicações para os dados, como por exemplo para a classificação de regiões, feito em [Silva et al. 2023], análise da mobilidade humana para planejamento urbano, assim como feito em [Santala et al. 2020], entre outras propostas de uso para os dados.

O trabalho está organizado da seguinte maneira, na Seção 2 são abordados os trabalhos que possuem maior relevância na motivação do trabalho e na metodologia escolhida. Na Seção 3 é discutida a metodologia com a descrição sobre a coleta de dados, ferramentas utilizadas e análise a cerca dos dados coletados. Na Seção 4 são abordadas as possíveis aplicações para os resultados do trabalho e por fim, na Seção 5 é apresentada a conclusão do trabalho e a proposição de trabalhos futuros.

¹<https://foursquare.com/>

2. Trabalhos Relacionados

Diante do atual contexto global, de alta disponibilidade de dados coletados de sensores sociais, diversos autores têm apresentado soluções envolvendo a mobilidade urbana. Nesta Seção são apresentados alguns trabalhos que fazem uso desses dados para o entendimento do meio urbano.

Em [Machado et al. 2018], os autores fazem uso de dados espaço-temporais extraídos de redes sociais *online*, buscando através de um estudo de caso na cidade de Nova York analisar as características sociais da cidade. A partir de um grafo social construído com base na proximidade física de usuários, os autores observam que essa proximidade pode indicar um encontro entre usuários. Também é analisada a distribuição de locais com base em sua categoria na cidade, comparando com o zoneamento oficial da cidade. Os autores concluíram que existem limitações relevantes no uso de dados das redes sociais *online*, como a dispersão de *check-ins* devido a ruídos presentes na base e a dependência no engajamento dos usuários em redes específicas.

Já em [Santos et al. 2020], os autores utilizam dados extraídos do *Twitter* buscando extrair a percepção dos usuários sobre áreas abertas das cidades de Chicago, Nova York e Londres a partir dos dados do *Twitter* e também de outros meios para coleta de *reviews* públicos através do *Google Places* e *Foursquare (Tips)*. A partir destes dados, os autores criaram um dicionário a fim de mapear as percepções a cerca de locais nas cidades analisadas, fazendo uso de técnicas de processamento de linguagem natural (PLN) para extrair a percepção humana para classificações das regiões estudadas.

Por fim, em [Senefonte et al. 2022], os autores propõem uma aplicação capaz de processar *check-ins* coletados a partir de *LBSNs* e prever padrões de mobilidade de turistas visitando outros países, com ou sem visitas anteriores registradas. Para isso foi necessário realizar a modelagem da mobilidade, extração do perfil dos usuários, para assim extrair a predição de mobilidade dos turistas. Ao final do trabalho, os autores concluem que os padrões disponíveis para serem explorados a partir das *LBSNs* são confiáveis para melhorar o entendimento a cerca do comportamento das pessoas como sensor social.

O presente trabalho se diferencia dos demais por apresentar de forma detalhada a coleta de dados para a modelagem e construção de uma base de dados utilizando *check-ins* em *LBSNs*, respeitando as políticas de privacidade atuais, impostas pelas próprias *LBSNs* e pelas leis em vigor em cada país. Assim, a partir desta análise, é possível a proposição de novas aplicações utilizando os dados coletados neste trabalho para diferentes domínios já estudados na literatura, como o agrupamento e classificação de regiões, propostos em [Silva et al. 2023, Le Falher et al. 2021, Zhang et al. 2021]. Aplicações voltadas à mobilidade urbana, propostas em [Santos et al. 2018, Machado et al. 2015, Brito et al. 2022, Silva et al. 2019]. Para o desenvolvimento do trabalho, foram consideradas técnicas de *Web Scraping* para obtenção de dados, de forma a respeitar a privacidade dos usuários, coletando apenas dados públicos e não sensíveis, possibilitando o seu uso em diversas aplicações.

3. Metodologia

É ilustrada, na Figura 1, a metodologia para a coleta e pré-processamento dos dados para que possa ser feita a análise dos mesmos. São discutidas, nas próximas subseções, todas as etapas da metodologia que foram utilizadas neste trabalho.

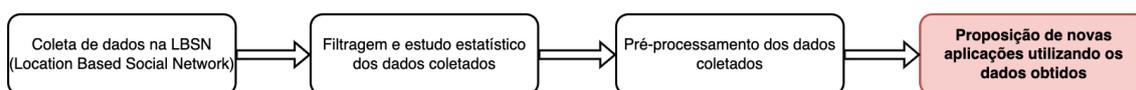


Figura 1. Fluxograma das Etapas do Trabalho.

3.1. Ferramentas

Para execução do presente trabalho, foi necessário a utilização de diferentes plataformas de desenvolvimento para a coleta e processamento dos dados. Primeiramente, foi utilizada a linguagem de programação *Python* para realizar requisições ao *Twitter API* ², cuja licença gratuita permite a coleta de 500.000 *tweets* por mês. As requisições foram feitas a partir de uma máquina virtual hospedada na *Microsoft Azure*. Após a coleta dos dados, *Python* também foi utilizado para o processo de extração dos dados coletados no *Twitter* e para a análise dos dados foram utilizadas as plataformas *Jupyter Notebook* e *Google Colab*. Todos os códigos utilizados no trabalho estão disponíveis em um repositório no *GitHub* ³.

3.2. Coleta de dados

De acordo com a política de privacidade do *Foursquare Swarm*, os *check-ins* são considerados como informação privada, o que impossibilita a coleta destes dados a partir da *Foursquare API*, que não permite tal ação. Como meio de contornar as limitações e coletar os dados provenientes do *Foursquare* é necessário buscar os dados a partir do *Twitter*, tendo em vista que existe a funcionalidade dentro do próprio *Swarm* que deixa em aberto a possibilidade de publicação do *check-in* em diferentes redes sociais, fazendo com que os *check-ins* se tornem públicos por opção dos usuários. A partir da obtenção dos *check-ins* pela *Twitter API*, foram coletadas aproximadamente 2,7 milhões de instâncias entre Maio de 2022 e Janeiro de 2023 ao redor do mundo inteiro, possibilitando uma análise mais abrangente sobre o comportamento das pessoas em cada país.

JSON 1 Exemplo de retorno do *Twitter API*.

```

{
  "data":{
    "id":"1560008545080430600",
    "text":"I'm at Avenida Brasil in Belo Horizonte, MG https://t.co/5FBbhBurSi",
    "timestamp":"2022-08-17 21:00:01.327915"
  },
  "matching_rules":[
    {
      "id":"1559949275353825281",
      "tag":"swarm"
    }
  ]
}
  
```

Em JSON 1 é representado um exemplo do retorno obtido através do *Twitter API*, onde é retornado o id do *tweet* no *Twitter*, o texto publicado, onde também se encontra

²<https://developer.twitter.com/>

³https://github.com/joaoaugustoss/Coleta_Check-ins

a *URL* referente ao *check-in* na rede *Foursquare Swarm*. Também é obtida a data e hora correspondentes ao horário de registro do *tweet* pelo usuário e por fim informações referentes às requisições feitas à *Twitter API*.

3.3. Pré-processamento

Após a coleta dos *tweets*, observou-se a presença de dados não úteis para o presente trabalho. Como mostra o exemplo ilustrado em JSON 1, dados como o id do *tweet* no *Twitter* e informações referentes à requisição para *Twitter API* não são úteis para a montagem da base de dados, e no campo texto publicado do usuário, foi feita a separação para que fosse armazenado apenas a *URL* contida neste texto. A partir da separação dos dados retornados pela *API*, a *URL* é aberta através de um código desenvolvido em *Python* para que sejam coletadas as informações necessárias do *check-in* dentro do código-fonte *HTML* referente à página do *check-in* no *Swarm*.

Atributo	Descrição do Atributo
<i>venueID</i>	Identificador da <i>venue</i> no <i>Foursquare</i>
<i>userID</i>	Identificador do usuário no <i>Swarm</i>
<i>venueName</i>	Nome da <i>venue</i> onde foi efetuado o <i>check-in</i>
<i>category</i>	Categoria da <i>venue</i> onde foi efetuado o <i>check-in</i>
<i>country</i>	País da <i>venue</i> onde foi efetuado o <i>check-in</i>
<i>city</i>	Cidade da <i>venue</i> onde foi efetuado o <i>check-in</i>
<i>timestamp</i>	Horário no qual o <i>check-in</i> foi compartilhado no <i>Twitter</i>
<i>latitude</i>	Latitude da <i>venue</i> onde foi efetuado o <i>check-in</i>
<i>longitude</i>	Longitude da <i>venue</i> onde foi efetuado o <i>check-in</i>

Tabela 1. Descrição da base de dados utilizada no trabalho.

Após o processamento dos dados, foi construída a base de dados, descrita na Tabela 1 com aproximadamente 2 milhões de instâncias, que após a exclusão de dados ausentes, restaram cerca de 1 milhão de instâncias para análise. Dentre as instâncias da base de dados, foram obtidos 985.661 *check-ins* em 234.164 *venues* únicas, em 15.519 cidades espalhadas em 148 países no mundo.

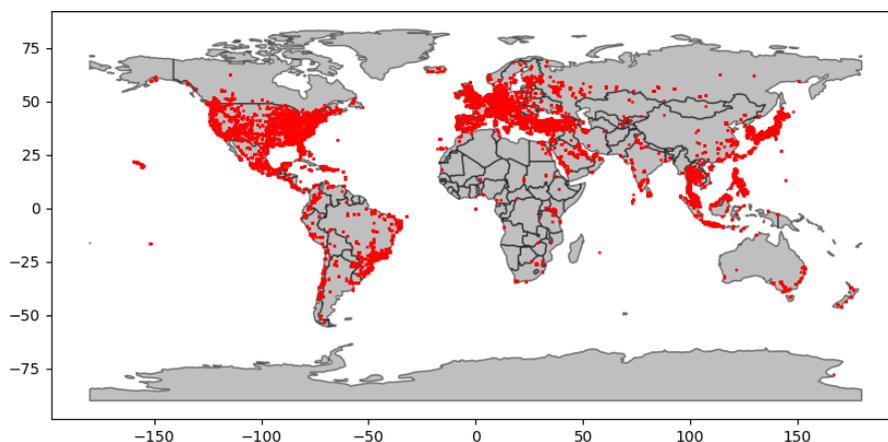


Figura 2. Distribuição dos *check-ins* coletados.

3.4. Análise

Após a coleta e o processamento dos dados, foi feita uma primeira análise com os dados coletados a partir do *Twitter*. Como mostra a Figura 2, os dados não estão distribuídos

de maneira uniforme pelo mundo, tendo focos principais no leste dos Estados Unidos, a região central da Europa e o Japão, que detém o maior número de *check-ins* na base de dados construída. Para tanto, a disparidade na disposição dos dados pode resultar em falhas nas análises feitas em várias cidades que possuem um número menor de registros dos usuários. Para isso, é desejável a combinação de bases coletadas de diferentes *LBSNs*, já que atualmente existem *LBSNs* mais internacionalizadas que o *Foursquare Swarm*, utilizado para as análises do presente trabalho. Análises mais detalhadas são encontradas em [Silva et al. 2023].

4. Aplicações

Com base nos dados coletados e descritos na Seção 3, e com a análise preliminar realizada na Seção 3.4, é possível a proposição de aplicações dos dados descritos neste trabalho nas seguintes vertentes:

Classificação: Agrupamento de regiões a fim de obter a classificação das mesmas, assim como proposto em [Silva et al. 2023].

Fluxo de Pessoas: Caracterização do fluxo de pessoas em diferentes meios de transporte dentro dos centros urbanos.

Fluxo de Veículos: Análise do trânsito para proposição de novas rotas evitando regiões que apresentam risco de segurança.

Planejamento: Análise de sentimento dos cidadãos e monitoramento e classificação de possíveis pontos de interesse dinamicamente para tomadas de decisão relacionadas ao planejamento urbano.

5. Conclusão e Trabalhos Futuros

Tendo em vista os dados apresentados na Seção 3 condizentes a *check-ins* em redes sociais, este trabalho apresentou a coleta de *check-ins* por meio de duas *LBSNs*, *Foursquare* e *Twitter*, e também descreveu a etapa de processamento dos dados, através de algoritmos desenvolvidos para deixar os dados aptos a serem objetos de estudo. Foi explorado o uso de ferramentas para análise dos dados a fim de entender a sua representação e relevância, apresentando como principal limitação, a dependência da interação do usuário com a *LBSN* para a obtenção dos dados. Com isso, tal análise leva a concluir que as redes sociais baseadas em geolocalização são importantes objetos de estudo para o sensoriamento social dentro do meio urbano, graças a crescente popularidade em todo o mundo e também pela crescente disponibilidade de dados de fácil acesso às empresas.

Para trabalhos futuros, torna-se fundamental explorar a coleta de dados em diferentes *LBSNs*, como por exemplo *Waze*, *Instagram*, *Facebook*, entre outras redes mais populares em mais regiões do mundo, além disto, estudar a implementação das aplicações propostas na Seção 4, gerando diferentes usos para os dados coletados neste trabalho.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – (PQ 310075/2019-0), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG – (Projeto PPM-00006-18 e PIBIC 2022/28009), e à Pontifícia Universidade Católica de Minas Gerais (PUC Minas) pelo suporte para o desenvolvimento deste trabalho.

Referências

- Brito, M., Santos, C., Oliveira, H., Cerqueira, E., and Rosário, D. (2022). Air pollution calculation for location based social networks multimodal routing service. In *Anais do VI Workshop de Computação Urbana*, pages 280–293. SBC.
- Gubert, F. R., Munaretto, A., and Silva, T. H. (2022). Multilayered analysis of urban mobility. In *Anais Estendidos do XXVIII Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 57–60. SBC.
- Le Falher, G., Gionis, A., and Mathioudakis, M. (2021). Where is the soho of rome? measures and algorithms for finding similar neighborhoods in cities. *Proceedings of the International AAAI Conference on Web and Social Media*, 9(1):228–237.
- Machado, K., Silva, T. H., de Melo, P. O. V., Cerqueira, E., and Loureiro, A. A. (2015). Urban mobility sensing analysis through a layered sensing approach. In *2015 IEEE International Conference on Mobile Services*, pages 306–312. IEEE.
- Machado, K. L., Boukerche, A., Cerqueira, E. C., and Loureiro, A. (2018). A data-centric approach for social and spatiotemporal sensing in smart cities. *IEEE Internet Computing*, 23(1):9–18.
- Santala, V., Costa, G., Gomes-Jr, L., Gadda, T., and Silva, T. H. (2020). On the potential of social media data in urban planning: Findings from the beer street in curitiba, brazil. *Planning Practice & Research*, 35(5):510–525.
- Santos, F. A., Silva, T. H., Loureiro, A. A., and Villas, L. A. (2020). Automatic extraction of urban outdoor perception from geolocated free texts. *Social Network Analysis and Mining*, 10:1–23.
- Santos, F. A., Silva, T. H., Loureiro, A. A. F., and Villas, L. A. (2018). Uncovering the perception of urban outdoor areas expressed in social media. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 120–127. IEEE.
- Senefonte, H. C. M., Delgado, M. R., Lüders, R., and Silva, T. H. (2022). Predictour: Predicting mobility patterns of tourists based on social media user’s profiles. *IEEE Access*, 10:9257–9270.
- Silva, J., Cunha, F., and Guimarães, S. (2023). Estudo do comportamento de consumo de bebida em centros urbanos usando redes de sensoriamento participativo. In *Anais do VII Workshop de Computação Urbana*, pages 68–81, Porto Alegre, RS, Brasil. SBC.
- Silva, T., De Melo, P. V., Almeida, J., Musolesi, M., and Loureiro, A. (2014). You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 466–475.
- Silva, T. H., Viana, A. C., Benevenuto, F., Villas, L., Salles, J., Loureiro, A., and Quercia, D. (2019). Urban computing leveraging location-based social network data: a survey. *ACM Computing Surveys (CSUR)*, 52(1):1–39.
- Zhang, M., Li, T., Li, Y., and Hui, P. (2021). Multi-view joint graph representation learning for urban region embedding. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 4431–4437.