

Modelagem do Interesse por Áreas Urbanas Usando Redes Sociais Baseadas em Localização

Gustavo H. Santos
Federal University of Technology - Parana
Curitiba, Brazil
gustavohenriquesantos@alunos.utfpr.edu.br

Myriam Delgado (Co-advisor)
Federal University of Technology - Parana
Curitiba, Brazil
myriamdelg@utfpr.edu.br

Thiago H. Silva (Advisor)
Federal University of Technology - Parana
Curitiba, Brazil
thiagoh@utfpr.edu.br

ABSTRACT

Location-Based Social Networks (LBSNs) can help model users' interests in urban areas in several ways. In the present work, we focus on Interest Networks (iNETs), which result from modeling LBSN data into graphs. The present study provides insights into which areas are frequently visited together by getting data from two distinct LBSNs, Foursquare and Google Places. Although the studied LBSNs differ in nature, with data varying in regularity and purpose, both modeled iNETs revealed similar urban behavior patterns and were likewise impacted by socioeconomic and geographic factors. Also, we discuss the development of a tool to empower urban studies and the by-products of this research.

KEYWORDS

LBSNs, grafos, modelagem comportamental, interesses, áreas urbanas

1 INTRODUÇÃO

Redes Sociais Baseadas em Localização (LBSNs, do inglês *Location-Based Social Networks*) são uma ferramenta amplamente usada na análise dos fenômenos urbanos [3, 4, 9, 14, 16, 17, 20, 23, 25]. Como as LBSNs fornecem dados que contêm, entre outras coisas, as preferências dos usuários por locais específicos de uma cidade [19, 22], uma modelagem apropriada pode aprimorar a compreensão da conduta destes usuários em seus ambientes urbanos.

O trabalho de pesquisa descrito neste artigo tem como objetivo geral entender parte do comportamento de usuários em áreas urbanas, usando dados de LBSNs. Como objetivos específicos, busca-se propor a modelagem dessa dinâmica social através de grafos - as Redes de Interesse; agregar fatores socioeconômicos, políticos e raciais à análise e por fim trabalhar diferentes granularidades na definição das regiões.

Em uma primeira etapa da pesquisa, fez-se uma coleta dos dados de duas LBSNs, Google Places e Foursquare, na cidade de Curitiba. O Google Places é uma plataforma na qual os usuários podem fazer avaliações sobre os locais que eles frequentaram. Já o Foursquare é uma rede social na qual as pessoas compartilham sua localização através de *check-ins*. Com esses dados geolocalizados, faz-se uma agregação das avaliações em um grafo não direcionado (rede de interesse) no qual os nós representam bairros da cidade e arestas representam o interesse dos usuários entre os bairros. Esses resultados evidenciam que, apesar de diferenças no uso e no propósito das plataformas, as duas LBSNs apresentam padrões similares em relação ao comportamento dos usuários no ambiente urbano. Os

resultados deste trabalho sugerem ainda que, entre diversos fatores analisados, a distância entre regiões urbanas influencia o comportamento das pessoas. Ainda na primeira etapa, foram incorporados às redes de interesse, dados socioeconômicos e preferências políticas que caracterizam os bairros analisados, como forma de avaliar sua influência nas escolhas dos usuários. Na segunda etapa da pesquisa, de modo a padronizar essas análises e expandir, mais facilmente, esse estudo para regiões do mundo com menos dados disponíveis, foi desenvolvida uma ferramenta que fornece uma subdivisão em hexágonos de tamanhos ajustáveis para grande parte das cidades do mundo.

Vale destacar o impacto da pesquisa realizada, que resultou em 3 trabalhos aceitos [5, 6, 15], sendo 2 em workshop de evento (qualis A4), um em evento internacional (qualis A2) e 2 convites para extensão em revista internacional (Qualis A2, JCR 3.5). O primeiro trabalho [15] foi resultado direto da primeira etapa e teve sua relevância atestada por um convite para extensão no *Journal of Internet Services and Applications* (JISA). Já o segundo e terceiro trabalhos representam contribuições indiretas; o segundo [5] foi premiado em 3o lugar no Courb 2024, também convidado para extensão no JISA e o terceiro trabalho [6] foi aceito para o Asonam 2024.

2 TRABALHOS RELACIONADOS

Dado que o comportamento urbano é uma parte fundamental do funcionamento de cidades, muitos estudos recentes ressaltam o uso de LBSNs para as suas análises. Numa avaliação comparativa entre várias LBSNs, tanto os benefícios das plataformas para o estudo dos fenômenos urbanos quanto os desafios envolvidos foram reconhecidos em [11]. Já [12] evidencia os diversos cenários em que os dados são úteis, como, por exemplo, na compreensão da dinâmica social, em intervenções urbanas específicas, no turismo, no papel da infraestrutura verde, entre outros. Buscando entender se diferentes LBSNs podem ser utilizadas juntas para o entendimento urbano, [21] usa o Instagram e o Foursquare para mostrar que um *check-in* do Foursquare pode trazer informação sobre a categoria de um estabelecimento comentado em uma publicação do Instagram.

Outros trabalhos buscam relacionar diferentes fatores com a mobilidade. Com dados geolocalizados do Twitter, [1] explora os relacionamentos entre as distâncias percorridas pelos usuários, as regiões que eles frequentam, a popularidade na plataforma e a linguagem usada nas postagens, identificado um padrão de mobilidade no qual distâncias menores têm maior preferência do que distâncias maiores. Em outra linha de investigação, [24] avalia se dados do Foursquare se assemelham aos dados informados pela WTO (World Tourism Organization), na questão do turismo internacional.

Diferente dessas abordagens, esta pesquisa tem como objetivo usar LBSNs para entender os fatores que influenciam o interesse

In: IV Concurso de Trabalhos de Iniciação Científica (CTIC 2024). Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web (CTIC'2024). Juiz de Fora/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2024.
© 2024 SBC – Sociedade Brasileira de Computação.
ISSN 2596-1683

Considerações éticas: Este trabalho não envolve pesquisas com seres humanos. Todas as informações sensíveis, por exemplo, os nomes de usuários, foram anonimizadas previamente para garantir a privacidade dos usuários. Além disso, são utilizados dados públicos disponíveis na Web.

4 RESULTADOS

Etapa 1: Ao construir as redes de Interesse para ambas as LBSNs, na cidade de Curitiba, primeiramente analisou-se a similaridade entre as duas através da correlação de Pearson. Foram comparados os pesos das arestas das duas redes obtendo uma correlação de 0,875, observando-se assim uma forte indicação da similaridade dessas redes de interesse. Para investigar se os bairros mais importantes eram os mesmos para as duas redes de interesse, empregou-se a correlação de Kendall's Tau. Para isso, construiu-se um *ranking* dos bairros, usando a centralidade por autovetor que indica a importância do nó na rede, comparou-se esse *ranking* entre as duas redes obtendo uma correlação de 0,646. Essa comparação pode ser vista na Figura 3 para os bairros mais importantes de cada rede. Dessa forma, observou-se uma conexão moderada entre as características de importância dos bairros para as duas LBSNs.

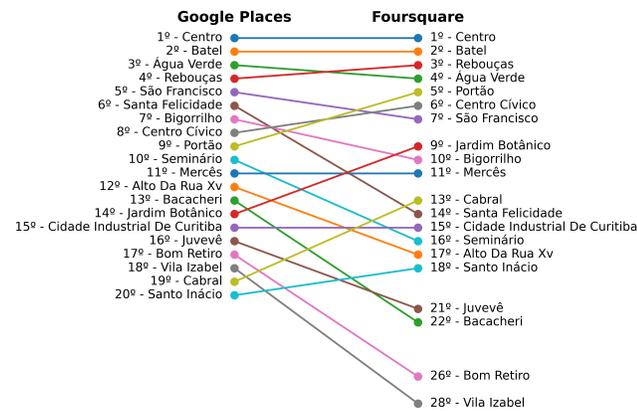


Figura 3: Comparação dos nós mais importantes de cada rede

Após calcular a diferença entre os bairros de Curitiba para cada um dos fatores (socioeconômicos, políticos e raciais), estas diferenças foram associadas aos pesos das arestas para ambas as Redes de Interesse (Google Places e Foursquare). Esse processo foi feito tanto para a rede completa, como para uma rede filtrada, selecionado apenas as arestas com peso ≥ 5 , de modo a capturar áreas conectadas por interesse mais elevado. A correlação de Spearman foi usada para entender quais desses fatores estão mais associados com o interesse dos usuários. Foi obtida uma correlação inferior a 0,25 em todos os fatores analisados, exceto a distância geográfica, que apresentou correlações entre -0,3 e -0,55 nos cenários considerados, ambas com $p \leq 10^{-4}$. Isso indica que os fatores de renda média mensal, composição racial e polarização política não ajudam a explicar o interesse dos usuários pelas áreas urbanas, logo as pessoas não costumam frequentar, necessariamente, lugares que sejam similares em renda, raça, ou opiniões políticas, quando seus interesses são modelados por essas LBSNs. Entretanto, a distância geográfica

teve um impacto mais relevante, mesmo que moderado, indicando a preferência dos usuários por regiões que são mais próximas entre si.

Etapa 2: Como forma de ampliar a análise da etapa 1 para diferentes granularidades, foi desenvolvida a h3-cities cujos resultados ilustrativos estão na Figura 4. Nela, há uma comparação entre os bairros de Curitiba e uma subdivisão feita pela ferramenta, com a resolução 8, ou seja, os hexágonos têm uma área média de 0,737 km². Com essa ferramenta, é possível dividir os bairros em regiões menores e verificar se as redes de interesse formadas pelas diferentes LBSNs continuam a apresentar resultados semelhantes. Outra questão importante ao trabalhar com uma subdivisão customizada de cenários urbanos é que dados de terceiros, como o IBGE, não fornecem dados especificamente para essa subdivisão. Isso que pode ser resolvido usando uma aproximação com dados dos bairros, ou com a exploração de outras características advindas de datasets geolocalizados, como, por exemplo, as categorias de estabelecimentos presentes naquela região, para a análise urbana.

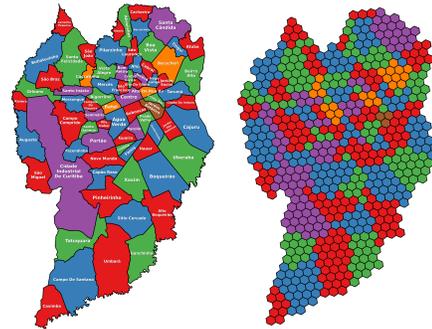


Figura 4: Comparação da segmentação de Curitiba através de bairros e uma subdivisão via h3-cities (resolução 8).

Em termos de produtos gerados pela pesquisa, além da ferramenta descrita anteriormente, destacam-se os 3 trabalhos aceitos para eventos (dois nacionais e um internacional). Os resultados diretos da etapa 1, desenvolvidos pelo aluno de IC, foram publicados em [15], cujo impacto foi atestado por um convite para extensão no periódico JISA, extensão esta que deverá englobar os resultados diretos da etapa 2. Entretanto, a pesquisa resultou também em resultados indiretos, através de contribuições na preparação de dados, construção de figuras, e desenvolvimento de parte dos resultados, que possibilitaram a publicação de outros 2 trabalhos em eventos [5, 6], os quais exploram o conceito de assinaturas culturais, através da Teoria Scenes, que com dados provenientes de LBSNs, pode, por exemplo, tentar expressar as características culturais de uma dada região. Em [5], o trabalho explora o processo de agrupamento de bairros com características culturais (assinatura cultural) semelhantes na cidade de Curitiba. Este trabalho foi premiado com o 3o lugar no workshop Courb 2024 e, assim como [15], também recebeu convite para extensão no JISA. Em [6], o trabalho explora a transferência de conhecimento de uma LBSN (Yelps) para outra (Google Places) na formação das assinaturas culturais, explorando dados de diferentes cidades ao redor do mundo, e também de todos

os estados americanos. Como forma de validação dos resultados para os EUA, os dados foram comparados com aqueles obtidos via American Values Survey (AVS), um resultado desenvolvido pelo aluno de IC.

5 CONCLUSÃO

Esse artigo apresentou um resumo do estudo desenvolvido sobre a modelagem do interesse de usuários de redes sociais por diferentes áreas urbanas. Para isso, foi realizada na primeira etapa da pesquisa uma investigação na cidade de Curitiba usando redes de interesse, i.e. *Location-Based Social Networks* (LBSNs) modeladas em grafos. Nos experimentos observou-se similaridade na modelagem do comportamento urbano, a partir de duas fontes de dados diferentes, Google Places e Foursquare, em particular quando se consideram as arestas, e os nós mais centrais das respectivas redes de interesse.

Ainda na primeira etapa da pesquisa, investigou-se a possibilidade de entender as escolhas dos usuários de LBSNs, em relação ao seu interesse por regiões urbanas, usando características socioeconômicas, de polarização política e distância geográfica das áreas frequentadas por esses usuários. Com isso, concluiu-se que a distância geográfica foi o único fator associado com o interesse dos usuários, mesmo que de forma moderada. Ou seja, regiões próximas costumam ser mais frequentadas pelos mesmos usuários do que regiões mais distantes.

Na última etapa que ainda se encontra em andamento, com o desenvolvimento da h3-cities, se torna possível a construção de redes de interesse em diversas cidades. Por exemplo, obtendo-se dados de uma LBSN, como os *datasets* públicos usados para este trabalho, pode-se investigar o interesse dos usuários, sem precisar buscar pelas subdivisões de uma cidade, o que pode ser difícil dependendo da região estudada. Ao segmentar qualquer cidade, que tenha dados disponíveis no OpenStreetMap, permite-se averiguar como o comportamento dos usuários varia sob diferentes granularidades. Logo, essa ferramenta alavanca um estudo padronizado para o cenário urbano.

Para trabalhos futuros, prevê-se a possibilidade de investigar outras cidades, datasets de outros períodos, além de explorar os efeitos provocados por diferentes granularidades na definição das áreas urbanas e validar os resultados aqui apresentados, desenvolvidos com uma quantidade de dados pequena, em datasets de maior escala. Dessa forma, seria possível analisar como os dados de diferentes LBSNs são influenciados pelo espaço temporal e escalas geográficas na modelagem do comportamento urbano, e ainda inferir quais características ajudam a explicar o interesse dos usuários pelas diferentes áreas urbanas.

6 AGRADECIMENTOS

Esta pesquisa foi apoiada pelo trabalho da mestranda Fernanda Gubert, pelo projeto FAPESP SocialNet (2023/00148-0) e CNPq (313122/2023-7, 314603/2023-9 e 441444/2023-7).

REFERÊNCIAS

- [1] Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Sui. 2021. Exploring Millions of Footprints in Location Sharing Services. *Proc. of ICWSM* 5, 1 (Aug. 2021), 81–88.
- [2] Kelsey Jordahl et al. 2020. *geopandas/geopandas: v0.8.1*.
- [3] Ana P.G. Ferreira, Thiago H. Silva, and Antonio A.F. Loureiro. 2020. Uncovering spatiotemporal and semantic aspects of tourists mobility using social sensing. *Computer Communications* 160 (2020), 240–252.
- [4] Ana Paula G. Ferreira, Thiago Henrique Silva, and Antonio Alfredo Ferreira Loureiro. 2015. Beyond Sights: Large Scale Study of Tourists' Behavior Using Foursquare Data. In *Proc. of ICDMW*. 1117–1124.
- [5] Fernanda Gubert, Gustavo H. Santos, Myriam Delgado, Daniel Silver, and Thiago H Silva. 2024. Criação de Assinatura Cultural de Áreas Urbanas com Estabelecimentos Geolocalizados na Web. In *Proc. of Courb*. Niteroi, RJ, 1 – 14.
- [6] Fernanda R Gubert, Gustavo H. Santos, Myriam R. B. S. Delgado, Daniel Silver, and Thiago H Silva. 2024 (to appear). Culture Fingerprint: Identification of Culturally Similar Urban Areas Using Google Places Datas. In *Proc. of Asonam*. Cosenza, Italia, 1 – 11.
- [7] Ruining He, Wang-Cheng Kang, and Julian McAuley. 2017. Translation-based Recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. ACM.
- [8] Peng Huang and Carter T. Butts. 2023. Rooted America: Immobility and Segregation of the Intercounty Migration Network. *American Sociological Review* 88, 6 (Nov. 2023), 1031–1065.
- [9] Lucas Ladeira, Allan Souza, Geraldo Rocha Filho, Thiago Henrique Silva, and Leandro Villas. 2019. Serviço de Sugestão de Rotas Seguras para Veículos. In *Anais do XXXVII SBRC* (Gramado). SBC, Porto Alegre, RS, Brasil, 608–621.
- [10] Xi Liu, Clio Andris, and Bruce A. Desmarais. 2019. Migration and political polarization in the U.S.: An analysis of the county-level migration network. *PLOS ONE* 14, 11 (Nov. 2019), e0225405.
- [11] Pablo Martí, Leticia Serrano-Estrada, and Almudena Nolasco-Cirugeda. 2019. Social media data: Challenges, opportunities and limitations in urban studies. *Comp., Envir. and Urban Systems* 74 (2019), 161–174.
- [12] Almudena Nolasco-Cirugeda and Clara García-Mayor. 2022. Social dynamics in cities: Analysis through LBSN data. *Procedia Computer Science* 207 (2022), 877–886.
- [13] Rajiv Pasricha and Julian McAuley. 2018. Translation-based factorization machines for sequential recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. ACM.
- [14] Ville Santala, Sandro Miczevski, Saulo A de Brito, Ariane Lao Baldykowski, Tatiana Gadda, Nadia Kozievitch, and Thiago H Silva. 2017. Making Sense of the City: Exploring the Use of Social Media Data for Urban Planning and Place Branding. In *Anais do I COURB* (Belém). SBC, Porto Alegre, RS, Brasil.
- [15] Gustavo H. Santos, Fernanda Gubert, Myriam Delgado, and Thiago H Silva. 2024. Redes de Interesse: comparando o Google Places e Foursquare na captura da escolha de usuários por áreas urbanas. In *Proc. of Courb*. Niteroi, RJ, 1 – 14.
- [16] Helen C. Mattos Senefonte, Myriam Regattieri Delgado, Ricardo Lüders, and Thiago H. Silva. 2022. PredicTour: Predicting Mobility Patterns of Tourists Based on Social Media User's Profiles. *IEEE Access* 10 (2022), 9257–9270.
- [17] Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, and Antonio A. F. Loureiro. 2017. Uma Fotografia do Instagram: Caracterização e Aplicação. *Revista Brasileira de Redes de Computadores e Sistemas Distribuídos* (2017).
- [18] Thiago H. Silva, Pedro Vaz de Melo, Jussara Almeida, Mirco Musolesi, and Antonio Loureiro. 2017. A large-scale study of cultural differences using urban data about eating and drinking preferences. *Information Systems* 72, Supplement C (2017), 95 – 116.
- [19] Thiago H Silva and Antonio AF Loureiro. 2016. Computação urbana: Técnicas para o estudo de sociedades com redes de sensoriamento participativo. *Sociedade Brasileira de Computação* (2016).
- [20] Thiago H Silva and Daniel Silver. 0. Using graph neural networks to predict local culture. *Environment and Planning B: Urban Analytics and City Science* 0, 0 (0).
- [21] Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, Juliana Salles, and Antonio A. F. Loureiro. 2013. A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior. In *Proc. ACM SIGKDD UrbComp '13*. Chicago, USA.
- [22] Thiago H. Silva, Aline Carneiro Viana, Fabrício Benevenuto, Leandro Villas, Juliana Salles, Antonio Loureiro, and Daniele Quercia. 2019. Urban Computing Leveraging Location-Based Social Network Data: A Survey. *Comput. Surveys* 52, 1 (Feb. 2019), 1–39.
- [23] Daniel Silver and Thiago H. Silva. 2023. Complex causal structures of neighbourhood change: Evidence from a functionalist model and yelp data. *Cities* 133 (2023), 104130.
- [24] Lucas E.B. Skora, Helen C.M. Senefonte, Myriam Regattieri Delgado, Ricardo Lüders, and Thiago H. Silva. 2022. Comparing global tourism flows measured by official census and social sensing. *Online Social Networks and Media* 29 (2022), 100204.
- [25] David A. M. Veiga, Gabriel B. Frizzo, and Thiago H. Silva. 2019. Cross-cultural study of tourists mobility using social media. In *Proc. of WebMedia* (Rio de Janeiro, Brazil) (*WebMedia '19*). ACM, 313–316.