

Estudo intercultural da mobilidade de turistas utilizando mídia social

Cross-cultural study of tourist mobility using social media

RESUMO

A grande disponibilidade de dados gerados por usuários em redes sociais, principalmente naquelas baseadas em localização, permite estudar o comportamento de sociedades urbanas em uma escala sem precedentes. Utilizando tais dados, particularmente, explorando check-ins do Foursquare-Swarm e avaliações feitas no TripAdvisor, este trabalho propõe estudar a mobilidade de turistas e residentes em cidades distintas ao redor do mundo. Demonstrando a aplicabilidade de dados gerados por redes sociais baseadas em localização, construímos um modelo baseado em grafos para representar a mobilidade dos dois grupos de usuários estudados em diferentes períodos do dia. Essa representação captura, de certa forma, a semântica da mobilidade, levando em consideração não somente as coordenadas geográficas, mas também o tipo de estabelecimento visitado. Os resultados indicam que residentes, independente da cidade, possuem padrões de mobilidade mais similares entre si que quando comparados com turistas. Observamos ainda que a proximidade geográfica, como em casos de cidades do mesmo país, influencia a similaridade observada nos dois grupos. Nossos resultados sinalizam que podem ser desenvolvidos novos sistemas de recomendação específicos para turistas, que considerariam a origem do usuário.

PALAVRAS-CHAVE: Mídia social. Turistas. Mineração de dados (Computação).

ABSTRACT

The vast availability of user-generated data in social networks, mainly in those based on location, enables the study of urban societies in an unprecedented scale. By utilizing such data, particularly, exploring check-ins from Foursquare-Swarm and reviews made in TripAdvisor, this work proposes the study of the mobility of tourists and residents in various cities around the world. Showcasing the inherent applicability for data generated by location based social networks, we built a model based on graphs to represent the mobility of both groups of studied users under distinct periods of the day. This representation demonstrates, in a certain way, the mobility's semantic, assuming not only the geographic coordinates but also the visited venue's type. The results indicate that residents, despite being from different cities, possess mobility patterns more similar between themselves, than when compared with tourists. We also observed that the geographical proximity, as in cases with two cities in the same country, influences the observed similarity between both groups. Our results suggest viability towards the development of new recommendation systems specific for tourists, which would regard the user's origin.

KEYWORDS: Social media. Tourists. Data mining.

Gabriel Branco Frizzo
gfrizzo@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil

Thiago Henrique Silva
thiagoh@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil

David Augusto Marcelino Veiga
davidaug23.7@gmail.com
Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brasil

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autorial: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



INTRODUÇÃO

A grande disponibilidade de dados gerados por usuários nas redes sociais baseadas em localização (LBSN), ou seja, um tipo especial de rede social online que leva em consideração aspectos geográficos em serviços diversos, permite a realização de estudos sobre sociedades urbanas em uma escala sem precedentes. Esses estudos permitem a obtenção de *insights* que podem ser utilizados em diversos setores, como no setor turístico.

A atividade turística é importante por diversos aspectos, por exemplo, o econômico. Em 2017, o turismo no mundo considerando somente turistas estrangeiros gerou uma receita de US\$1,340 bilhão, referente à chegada de 1,3 milhão de turistas internacionais. Além disso, um a cada dez empregos é voltado ao turismo, conforme dados da *World Tourism Organization*(2018). Para que a atividade turística de uma certa localidade seja sustentável, um dos fatores primordiais é oferecer um bom serviço. A personalização de atividades é um caminho importante para esse objetivo.

Nessa direção, uma questão central deste trabalho é: turistas que compartilham a mesma cultura tendem a realizar atividades mais similares? Para estudar essa questão, consideramos a mobilidade de turistas e residentes em diferentes localidades de países distintos. Para isso, exploramos *check-ins* do *Foursquare-Swarm* (de dois datasets distintos). A partir desses dados, construímos um modelo baseado em grafos para representar a mobilidade de turistas e residentes em diferentes períodos do dia. Essa representação captura, de certa forma, a semântica da mobilidade, utilizando não apenas as coordenadas geográficas, mas também o tipo de local visitado.

Ao descrever os grupos de usuários através da sua mobilidade associada a categorias de locais, foi possível realizar um agrupamento hierárquico. Os resultados indicam que os residentes e turistas, em geral, possuem padrões de mobilidade distintos. Possivelmente essa diferença está associada à rotina desenvolvida por residentes que, tipicamente, não é desenvolvida por turistas. Observamos também que a região geográfica, como o país, influencia a semelhança observada em ambos os grupos.

Esses resultados preliminares sinalizam que existe uma oportunidade para desenvolver sistemas de recomendação personalizados para turistas levando em consideração a origem do usuário.

MATERIAL E MÉTODOS

Neste trabalho utilizamos dois *datasets*, ambos extraídos da plataforma *Foursquare-Swarm*. O primeiro *dataset* denominado *Dataset1*) contém dados de *check-ins* realizados no período de Abril de 2017 a Abril de 2019 em Chicago, Istambul, Jacarta, Londres, Los Angeles, Nova Iorque, Paris, Seul, Singapura e Tóquio. Os dados foram fornecidos por um terceiro, de forma agregada, representando a quantidade de transições realizadas entre cada local único (id do local) dessas cidades. Além desses dados, cada id de local possui uma

classificação, conforme as dez categorias principais estabelecidas pelo Foursquare: Alimentação; Ar Livre e Recreação; Arte e Entretenimento; Compras e Serviços; Faculdade e Universidade; Evento; Profissional e Outros Locais; Residência; Viagem e Transporte; e Vida Noturna.

O segundo *dataset* (*Dataset2*) é composto de *check-ins* públicos feitos entre Setembro de 2010 e Junho de 2014. Os dados foram obtidos utilizando um *Web Crawler* desenvolvido especialmente para esse propósito. O *Dataset2* oferece uma informação mais refinada, já que os dados não estão agregados por local, e cada *check-in* é um registro diferente. Desta forma, é possível realizar análises mais informativas. No entanto, o volume de dados é reduzido, se comparado ao do *Dataset1*. Este *trade-off* dá maior flexibilidade aos estudos. Cada registro neste *dataset* consiste do id do usuário que fez a atividade, a localização GPS do local visitado, o horário e a categoria do local.

Utilizamos ainda dados do *TripAdvisor*. O *TripAdvisor* é uma rede social voltada para viajantes planejarem suas viagens. Esse *dataset* é constituído das avaliações dos locais feitas por viajantes (Miguéns, J, 2008). Construímos um coletor de dados que extrai, para cada cidade citada do *Dataset1*, as 20 atrações turísticas com mais avaliações.

Para a realização do estudo proposto, é fundamental identificar turistas e residentes. Para realizar esse processo no *Dataset1*, primeiramente identificamos as localidades do *dataset* que são turísticas (POI) utilizando, para isso, os dados do *TripAdvisor*. Em seguida identificamos todas as transições que acessam um POI, tanto como destino quanto como origem. Assumimos que todas essas transições são feitas por turistas. Todas as outras transições são consideradas de residentes. Já no *Dataset2*, a habilidade de individualizar usuários possibilitou deduzir seus locais de residência, ainda que com possíveis inconsistências. Para isso, foi escolhido o país no qual o usuário passa mais tempo durante o período analisado. No caso em que dois *check-ins* subsequentes são feitos em países diferentes, é admitido que o usuário passou esta janela de tempo no primeiro país. Esse método pode trazer imprecisões, mas dado o grande volume de dados estudados, é possível relevá-las.

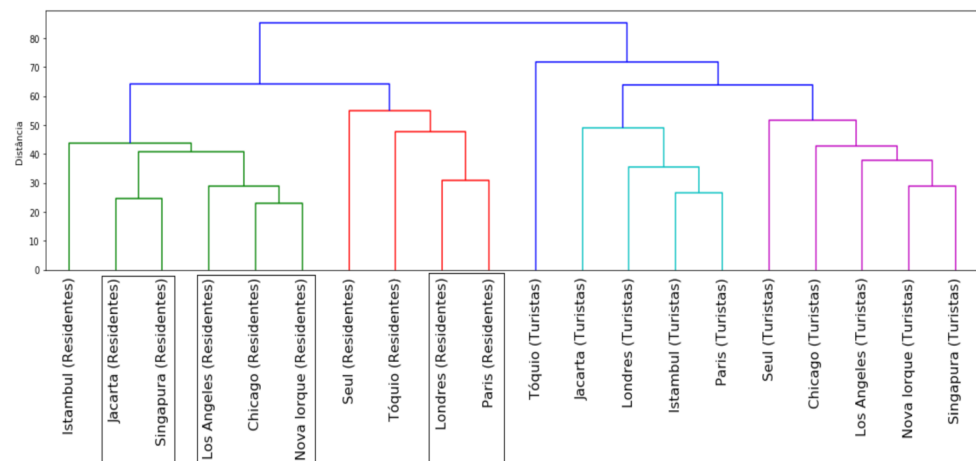
Foi gerado um grafo de transição bidirecional $G_i(V,E)$ para cada região i disponível no *dataset* (cidade para o *Dataset1*, país para o *Dataset2*), onde os vértices $v_i \in V$ representam as categorias do Foursquare-Swarm. Uma aresta $e(i,j)$ é formada caso exista uma transição do vértice v_i para o vértice v_j . Foi atribuído um peso $w(i,j)$ para cada aresta de acordo com a quantidade de transições realizadas de um vértice v_i para um vértice v_j , desde que essa atividade tenha ocorrido no mesmo período do dia (Manhã: inclui os horários entre 6h00m e 9h59m Meio-Dia: 10h00m e 14h59m; Tarde: 15h00m e 18h59m; Noite: 19h00m e 23h59m; Madrugada: 00h00m e 05h59m). Para cada cidade, temos um grafo para turistas e outro para residentes.

Para cada grafo gerado, utilizamos as suas matrizes de adjacência (matriz 10x10), onde linhas e colunas são representadas pelas 10 categorias indicadas previamente, e o valor de cada elemento corresponde à quantidade de transições de uma categoria para a outra. Essas matrizes representam, de certa maneira, o comportamento do grupo analisado, como demonstra Silva (2012). Cada matriz é transformada, então, em um vetor de 100 posições, através da concatenação das suas linhas.

Esses vetores, para ambos os datasets, são usados em uma clusterização hierárquica. Esse processo consiste em agrupar os objetos de acordo com suas similaridades. Neste caso, os agrupamentos indicariam semelhanças comportamentais entre os grupos analisados. Para a clusterização, foi utilizada a distância *Canberra*, com o método de ligação *Ward*.

RESULTADOS E DISCUSSÃO

Figura 1 - Dendrograma do agrupamento



Fonte: Autoria própria (2019).

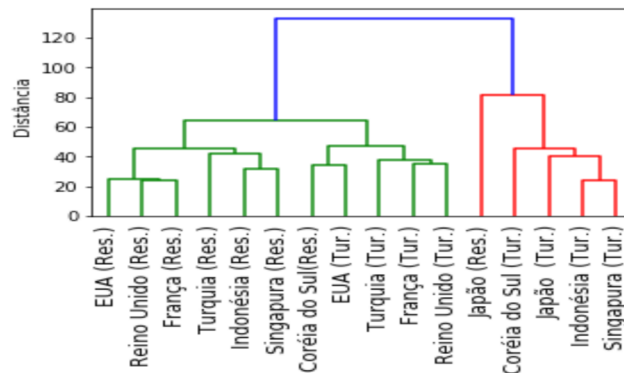
Os agrupamentos foram realizados em ambos os *datasets*. A Figura acima apresenta o agrupamento do *Dataset1* através de um dendrograma. Primeiramente, podemos observar que turistas estão em grupos significativamente separados dos residentes. Isso, de certa forma, é esperado, pois residentes tendem a realizar atividades relacionadas às suas rotinas diárias, que não são comuns a turistas.

Podemos constatar ainda que as transições classificadas como feitas por residentes apresentam maiores semelhanças entre cidades de uma mesma região. Foram encontrados três grupos relacionados com suas regiões geográficas: Estados Unidos (Los Angeles, Nova Iorque e Chicago), Europa (Londres e Paris), Sudeste Asiático (Jacarta e Singapura). As cidades da Ásia Oriental, Seul e Tóquio, não formam um grupo, mas estão próximas no espaço considerado. Istambul revela uma maior proximidade com as cidades do Sudeste Asiático e dos EUA, embora sutil. O resultado sugere correlação com as descobertas feitas por Brito et al.(2018), Yang et al.(2016) e Silva et al. (2014), onde cidades culturalmente próximas e, muitas das vezes, geograficamente próximas tendem a ser agrupadas.

Conforme indicado anteriormente, os *datasets* identificam turistas de forma diferente. Com base nos resultados de agrupamento observados para o *Dataset1*, temos a indicação de que a estratégia de identificação de turistas captura, no mínimo, parte do comportamento de turistas. No entanto, não é possível entender, somente com os resultados do *Dataset1*, o quão gerais são os resultados observados. Para estudar essa questão, realizamos uma nova análise com o *Dataset2*. Realizamos um agrupamento seguindo a mesma metodologia aplicada para o *Dataset1*. Essa análise visa ajudar a entender se POIs influenciam

significativamente na identificação de turistas. Como o *Dataset2* considera os países correspondentes às cidades do *Dataset1*, temos ainda a oportunidade de estudar se grandes cidades representam consideravelmente o comportamento observado para seus países.

Figura 2 – Dendrograma do agrupamento entre países do Dataset2



Fonte: Autoria própria (2019).

A Figura acima apresenta os resultados do agrupamento para o *Dataset2* através de um dendrograma. Assim como nos resultados do *Dataset1*, quase todos os agrupamentos de residentes e turistas ficaram bem separados. Os agrupamentos de residentes ficaram semelhantes aos da Figura 1, com exceção de japoneses e coreanos. Um fato interessante é que os residentes da Coreia do Sul parecem ser mais similares aos turistas ocidentais que aos residentes dos demais países. Um dos fatores que podem promover essa diferença é o grande número de turistas coreanos visitando países ocidentais. Podemos observar ainda que residentes do Japão possuem um comportamento distinto dos demais residentes, mas similar ao de turistas dos países asiáticos estudados. Esse fato pode retratar em parte a influência da cultura japonesa nos países asiáticos, além da grande quantidade de turistas japoneses nesses países. A proximidade da Turquia com a Indonésia (assim como Istanbul e Jacarta no *Dataset1*), pode estar relacionada a cultura religiosa compartilhada entre essas cidades, conforme o mapa cultural criado por Inglehart e Welzel na World Values Survey(2010), sendo então mais uma evidência de que o resultado é coerente com o aspecto estudado.

No geral, temos a indicação de que a forma de identificação de turistas não influencia significativamente os resultados. Assim, o resultado observado para o *Dataset1*, mais atual e maior, captura características importantes dos grupos estudados. Temos ainda a sugestão de que grandes cidades capturam o comportamento geral observado em seu país, embora essa estratégia deva ser utilizada com cautela, pois uma investigação mais profunda é demandada.

CONCLUSÃO

Nosso trabalho avança a literatura relacionada ao estudo de padrões culturais com dados de LBSNs. De posse de dois *datasets* distintos do *Fourquare-Swarm*, estudamos as diferenças de comportamento entre turistas e residentes em diferentes localidades, muitas com hábitos culturais distintos. Para isso,

modelamos o comportamento considerando a mobilidade de usuários associada a categorias de estabelecimentos. Os resultados sugerem que residentes de uma mesma região geográfica tendem a ter hábitos mais similares entre si. Observamos ainda que o comportamento estudado é diferente, no geral, entre turistas e residentes. Assim, a origem da residência do usuário pode influenciar nas suas escolhas, sendo possível explorar características culturais de cada nacionalidade em diferentes destinos. Ao levar isso em consideração, se torna clara a possibilidade de melhorar sistemas de recomendação de locais e serviços oferecidos para diferentes grupos de turistas.

Para trabalhos futuros, pretendemos aprofundar as análises das diferenças culturais observadas tanto do local de origem do usuário (nacionalidade), como do seu destino. Planejamos também estudar o impacto do tempo nas características analisadas, ou seja, do horário no qual as atividades são realizadas. Planejamos ainda verificar qual o impacto das estações do ano nas atividades dos usuários.

AGRADECIMENTOS

Agradecemos à UTFPR e ao CNPq pelo apoio ao projeto.

REFERÊNCIAS

Inglehart, R; Welzel, C. 2010. **Changing Mass Priorities: The Link between Modernization and Democracy.** Perspectives on Politics 8, 2 (June 2010), 551–567.

Miguéns, J; Baggio, R; Costa, C. 2008. **Social media and tourism destinations: TripAdvisor case study.** Adv. in tourism research 26, 28 (2008), 1–6

Saulo Brito, Ariane Baldykowski, Sandro Miczevski, and Thiago Silva. 2018. Cheers to Untappd! Preferences for Beer Reflect Cultural Differences Around the World. In Proc. of AMCIS.

Silva, Thiago H.; Vaz de Melo, Pedro O. S.; Almeida, Jussara M.; Musolesi, M; Loureiro, Antonio A. F. 2014. **You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare.** In Proc. of ICWSM Ann Arbor, USA

Silva, Thiago H.; Vaz de Melo, Pedro O. S.; Almeida, Jussara M.; Loureiro, Antonio A. F. 2012. **Visualizing the Invisible Image of Cities.** Proc. IEEE International Conference on Cyber, Physical and Social Computing

Yang, D; Zhang, D; Bingqing Qu. 2016. **Participatory Cultural Mapping Based on Collective Behavior Data in Location-Based Social Networks.** ACM Trans. Intell. Syst. Technol. 7, 3, Article 30 (Jan. 2016), 23 pages.