

Influências Regionais na Mobilidade de Turistas e Residentes Usando Dados de Mídia Social

David A. M. Veiga¹, Gabriel B. Frizzo¹, Helen C. M. Senefonte¹, Ricardo Lüders¹, Myriam R. B. S. Delgado¹, Thiago H. Silva^{1,2}

¹ Universidade Tecnológica Federal do Paraná
Curitiba, PR, Brasil

²University of Toronto
Toronto, ON, Canada

davidaug23.7@gmail.com, gfrizzo@alunos.utfpr.edu.br, helen@uel.br
{luders,myriam}@utfpr.edu.br, th.silva@utoronto.ca

Resumo. *As redes sociais baseadas em localização (LBSNs) disponibilizam uma nova variedade de possibilidades para obtenção de dados em larga escala, especialmente com o significativo aumento de usuários de mídias sociais. Nesse sentido, a proposta do trabalho atual é explorar dados de LBSNs para estudar as influências regionais no comportamento de turistas e residentes no âmbito de mobilidade, levando em consideração não somente as coordenadas geográficas, mas também o tipo de local visitado. Os resultados apresentados sugerem que residentes de uma mesma região geográfica tendem a ter hábitos mais similares entre si. Observou-se também que o comportamento entre turistas e residentes é diferente. Além disso, os experimentos indicaram que o país de origem de residência do usuário pode influenciar em suas escolhas enquanto turista em outros países. As análises realizadas permitiram observar também a existência de dois grupos distintos de turistas, sendo um deles mais previsível em termos de coordenadas geográficas e tipo de local visitado. A possibilidade de explorar as características culturais de cada nacionalidade em diferentes destinos mostra-se um caminho promissor para melhorar sistemas de recomendação de locais e serviços oferecidos para grupos de turistas distintos.*

1. Introdução

Estudos sobre sociedades urbanas utilizam diferentes fontes de dados em suas pesquisas. Dados de traços GPS, capturados, por exemplo, com o auxílio de dispositivos móveis, dados públicos disponibilizados por órgãos governamentais ou ainda dados online compartilhados em mídias sociais são algumas alternativas. Um tipo especial de mídia social que leva em consideração aspectos geográficos em serviços diversos são as redes sociais baseadas em localização (LBSNs). Considerando-se a vasta quantidade de dados disponibilizados pelas LBSNs, estas fornecem um novo e amplo espectro de possibilidades para a realização de estudos sobre sociedades urbanas em uma escala sem precedentes [Silva et al. 2019] que podem ser aplicados para solucionar problemas em diversas áreas.

Dados de LBSNs podem ser explorados para melhorar o entendimento da dinâmica de cidades, por exemplo, para detectar características sócio-econômicas em suas diferentes áreas geográficas [Huang and Wang 2016, Hristova et al. 2016],

facilitar o entendimento das fronteiras culturais e semelhanças entre sociedades [Gavilanes et al. 2014, Falher et al. 2015, Silva et al. 2017], detectar indicadores de características sobre vários aspectos de regiões da cidade, como, percepções de usuários e níveis de tráfego [Zheng et al. 2014, Ribeiro et al. 2014, Gu et al. 2014, Santala et al. 2017, Rodrigues et al. 2018, Santos et al. 2018] e também para auxiliar as atividades no setor turístico [Yang et al. 2016, Ferreira et al. 2019].

O setor turístico de uma cidade necessita de planejamento e possui importância em vários aspectos, especialmente o econômico. Considerando dados de turistas estrangeiros, o turismo mundial gerou uma receita de mais de um bilhão de dólares americanos, referente à chegada de 1,3 milhão de turistas internacionais [Organization 2018] apenas no ano de 2017. Outra estatística importante é em relação ao mercado de trabalho, um em cada dez empregos é voltado ao turismo, conforme dados da *World Tourism Organization* [Organization 2018]. Para que a atividade turística de uma localidade se mantenha atrativa, é primordial entender as preferências dos turistas de forma a oferecer serviços melhores e mais inteligentes.

Nesse sentido, o presente trabalho visa explorar dados de mídias sociais para estudar as influências regionais no comportamento de turistas e residentes no âmbito de mobilidade. As contribuições desse trabalho são variadas, indo de melhorias de sistemas de recomendações (via noção de distância comportamental) até melhor aproveitamento da capacidade de redes celulares (direcionando a demanda para outros serviços). O estudo desenvolvido pode ser usado em aplicações da área de cidades inteligentes, computação social, computação urbana e redes sociais. As questões centrais que norteiam essa pesquisa são: (i) Em termos de mobilidade, como os turistas se comportam nos países para os quais viajam? De forma similar aos residentes dos países de destino ou apresentam padrões de mobilidade semelhantes aos residentes dos seus próprios países de origem? (ii) Esses turistas tendem a ampliar sua área de visitação de forma exploratória ou se mantêm restritos a poucos pontos de maneira trivial? (iii) Turistas com o mesmo hábito cultural tendem a realizar atividades turísticas similares?

Para explorar essas questões, foi analisada a mobilidade de turistas e residentes em diferentes localidades de países distintos. Para isso, foram considerados *check-ins* do Foursquare-Swarm. Esses dados foram utilizados na construção de um modelo baseado em grafos que representa a mobilidade de turistas e residentes em momentos diferentes ao longo do dia. Essa representação capta a semântica da mobilidade, utilizando o tipo de local visitado além das coordenadas geográficas.

Primeiramente foi realizado um agrupamento hierárquico, permitindo a comparação de grupos de usuários através da descrição da mobilidade associada a categorias de locais. Os resultados mostram que os padrões de mobilidade associados a turistas e residentes tendem a ser bem distintos, diferença que é ocasionada muito provavelmente devido às suas diferentes rotinas.

Tendo por base o conceito de distância comportamental (uma inovação que representa uma importante contribuição deste trabalho), foram realizadas comparações entre diferentes métricas com o intuito de analisar a influência regional no comportamento de usuários. Os resultados apresentados mostram que os usuários possuem padrões de mobilidade diferentes e que as influências regionais têm efeitos distintos no comportamento de usuários em determinados países.

Outra contribuição do trabalho inclui o uso do cálculo de raio de giro para identificar e analisar a mobilidade de usuários de acordo com o fenômeno de *Returns* e *Explorers*, proposto por Pappalardo et al. (2015), mas dentro de um contexto até então não avaliado na literatura: *Residentes* × *Turistas*. Os resultados indicam que a maioria dos turistas avaliados neste trabalho pertencem à classe *Returns*, ou seja, o grau de previsibilidade das coordenadas geográficas e categorias visitadas é elevado, mas que essas medidas podem variar de acordo com a origem dos turistas e residentes considerados. Apesar da maioria dos turistas apresentarem um comportamento mais similar aos *Returns*, existe uma parte não desprezível de turistas que são categorizados como *Explorers*, e isso acontece para todas as nacionalidades estudadas.

Considerando-se as principais contribuições destacadas anteriormente, vale salientar que os resultados obtidos nesse trabalho ratificam que sistemas de recomendação de locais personalizados para turistas podem ser aprimorados, considerando informações adicionais como a origem do usuário e a sua característica cultural.

O restante do trabalho está organizado da seguinte maneira: A Seção 2 apresenta os principais trabalhos correlatos e mostra como o presente estudo contribui com novas formas de entender a mobilidade de turistas. A Seção 3 apresenta a metodologia e o conjunto de dados utilizados na pesquisa. A Seção 4 descreve os resultados obtidos. A Seção 5 faz uma discussão geral dos resultados. Por fim, na Seção 6 são apresentadas as considerações finais e direções para trabalhos futuros.

2. Trabalhos Correlatos

Diversas pesquisas na área de mobilidade de turistas utilizam dados extraídos de fontes tradicionais (questionários, entrevistas, etc). Por exemplo, Zieba (2017) examinou como características individuais de turistas austríacos influenciam na motivação de viagens e se seus hábitos em outros países são diferentes baseado em dados disponibilizados por questionários. Scuderi e Nogare (2018) analisaram padrões de consumo com o uso de cartão de turismo (que fornece descontos na cidade de Trentino) também explorando fontes tradicionais de dados.

Estudos que se baseiam em fontes de dados tradicionais geralmente sofrem com baixa escalabilidade, gerando trabalhos com estudos limitados, que exploram regiões menores, como uma cidade ou estado. O uso de dados LBSNs permite o aumento da escalabilidade para o estudo do comportamento social. Existem vários estudos nessa direção, incluindo estudos focados em aspectos culturais. Por exemplo, Silva et al. (2017) classificaram de maneira automática, as fronteiras culturais de usuários do Foursquare, baseadas em suas preferências de locais de *check-ins* na categoria alimentação (restaurantes, bares, etc). Já Brito et al. (2018) confirmam, através do uso de dados da rede social Untappd, que as preferências cervejeiras de usuários em diferentes locais ao redor do mundo também são características que descrevem seus hábitos culturais. Yang et al. (2016) ilustram a existência de agrupamentos entre cidades de regiões culturalmente próximas através do uso de dados de *check-ins* de residentes em diversas cidades. Vu et al. (2019) estudam as preferências de turistas malaios em diversos países e suas diferenças com turistas tailandeses. Os autores consideram apenas os dados de *check-ins* dos usuários. O estudo sugere que apesar da proximidade regional da Malásia e Tailândia, as nacionalidades devem ser levadas em conta para fins de publicidade e estratégia de mercado.

[Morais and Andrade 2014] apresentam um estudo de caso sobre a relevância de anotações geo-localizadas compartilhadas por turistas e residentes durante um evento turístico de grande porte. Os resultados do trabalho sugerem que a abordagem diferenciada dos dados compartilhados por residentes pode aumentar a habilidade dos participantes de encontrar informações relevantes. Long et al. (2013) também focam seus esforços na pesquisa do comportamento de turistas usando dados de LBSNs. Os autores estudam empiricamente os locais e seus respectivos check-ins baseado na informação da categoria dos locais. Eles exploram características temporais dos check-ins de turistas e aplicam o conceito de entropia de usuário para analisar suas movimentações. O trabalho também utiliza LDA (Latent Dirichlet Allocation) para gerar padrões de mobilidade dos turistas. Uma análise sobre as preferências de turistas e residentes foi feita por Ferreira et al. (2019). Utilizando grafos temporais, o trabalho sugere que cada grupo possui preferência de horário em determinados lugares.

A proposta do presente estudo se difere dos trabalhos correlatos supracitados devido ao uso e exploração da mobilidade dos usuários como uma característica descritiva. Essa característica foi utilizada para estudar a similaridade de grupos de turistas distintos e residentes em diferentes localidades em vários continentes. Em um estudo inicial anterior [Veiga et al. 2019], utilizou-se a mesma abordagem para avaliar dois conjuntos de dados distintos, para os quais foi possível concluir que grandes cidades capturam o comportamento geral observado em seu país. Além disso, ao testar duas formas diferentes de identificação de turistas, constatou-se que esse processo não influencia significativamente os resultados. Observou-se também, ainda que de forma preliminar, que a região de origem do usuário pode influenciar nas atividades realizadas.

O trabalho atual estende o estudo inicial de diferentes maneiras. A conclusão obtida em [Veiga et al. 2019], que grandes cidades capturam o comportamento geral observado em seu país, serviu de base para a análise da Seção 4.3, que filtra os dados para grandes cidades. Além disso, a conclusão anterior de que a origem do usuário influencia suas atividades foi ampliada ao considerar-se no presente trabalho que o destino também é relevante. Neste sentido, a análise feita por turistas de um país visitando outro país foi uma das principais inovações do presente trabalho. No estudo atual procura-se entender melhor a influência da origem do usuário em suas decisões relativas à mobilidade. Assim, realizou-se uma comparação mais abrangente entre turistas e residentes, obtendo-se diferentes grupos em relação aos seus padrões de mobilidade. Além disso, foi proposta uma abordagem para quantificar a influência regional na mobilidade baseada no conceito de distância comportamental. Por fim, foram estudados os padrões de mobilidade de turistas e residentes com relação ao fenômeno de *Returners* e *Explorers*.

3. Metodologia

Esta seção apresenta o conjunto de dados e os passos metodológicos utilizados neste estudo.

3.1. Conjuntos de Dados

O conjunto de dados utilizado (*Dataset*) 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. Este *Dataset* oferece uma informação refinada, sendo que cada registro consiste no *id* do usuário que fez a atividade, a localização

GPS do local visitado, o horário e a categoria do local (os nomes e organização são estabelecidos pelo Foursquare¹).

Objetivando maior volume e diversidade de dados, os seguintes países foram estudados, sem perda de generalidade do método: Alemanha, Argentina, Bélgica, Brasil, Canadá, Chile, Estados Unidos, Filipinas, Indonésia, Japão, Malásia, México, Reino Unido, Tailândia, Turquia.

3.2. Separação de Turistas e Residentes

Para a realização do estudo proposto, é fundamental identificar turistas e residentes. A capacidade de individualizar usuários no *Dataset* possibilitou deduzir seus locais de residência, mesmo com possíveis inconsistências. Para isso, foi escolhido como residência o país no qual o usuário passa a maior parte do tempo durante o período analisado. Em outros locais, ele é considerado turista. 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. Apesar de trazer algumas imprecisões, acredita-se que o impacto dessas premissas possa ser desconsiderado dado o grande volume de dados estudados. Trabalhos anteriores da literatura utilizaram com sucesso estratégias similares [Paldino et al. 2015, Ferreira et al. 2019].

3.3. Grafos de Mobilidade

Para a geração dos grafos de mobilidade, primeiramente, a identificação dos locais (id) no *Dataset* foi substituída por sua respectiva categoria. Foram utilizadas 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.

Em seguida, um grafo de transição bidirecional $G_m(V, E)$ foi obtido para cada país m disponível no *Dataset*, onde os vértices $v_i \in V$ representam cada uma das categorias do Foursquare-Swarm. Uma aresta $e(i, j)$ é formada caso um usuário tenha feito dois *check-ins* subsequentes (transição) do vértice v_i para o vértice v_j , dentro de um tempo limitado. Caso a janela de tempo entre os *check-ins* seja de mais de dois períodos do dia (definidos abaixo), não é considerada uma transição. Os períodos do dia incluem: (i) manhã entre 6:00-9:59; (ii) meio-dia entre 10:00-14:59; (iii) tarde entre 15:00-18:59; (iv) noite entre 19:00-23:59; (v) madrugada entre 00:00 e 05:59. Para cada país analisado, há um grafo de mobilidade para os residentes deste país e diversos grafos para seus 'emigrantes', representando seus turistas visitando cada um dos demais países.

3.4. Agrupamento

Cada grafo gerado é representado por sua matriz de adjacência (matriz 10×10), cujas linhas e colunas são as 10 categorias listadas na subseção 3.3 e o valor de cada elemento da matriz corresponde à quantidade de transições de uma categoria para a outra. As linhas da matriz são então concatenadas para formar um vetor de 100 posições denominado aqui *vetor de mobilidade*, estrutura que caracteriza, de certa maneira, as preferências de visitação do grupo. Desta forma, o comportamento é metrificado em um espaço 100-dimensional, para que grupos diferentes possam ser comparados posteriormente.

¹<https://developer.foursquare.com/docs/resources/categories>.

Em [Veiga et al. 2019] foi proposto um método que consiste em agrupar hierarquicamente os objetos de acordo com suas similaridades. No presente trabalho, a mesma metodologia foi considerada para a obtenção e análise de novos resultados. Foram testadas as combinações das distâncias *Euclidiana* e *Canberra*, com os métodos de ligação *Ward* e *Complete*, sendo que os resultados mais coerentes foram obtidos pela métrica *Canberra* combinada com o método *Ward*. Assim os resultados são apresentados somente para essa combinação.

3.5. Influência Regional no Comportamento de Turistas

Com o objetivo de mensurar os efeitos de cada país no comportamento de seus turistas, foi estabelecida uma comparação entre turistas e residentes de dois países A e B , conforme a Tabela 1. A comparação é feita considerando os vetores de mobilidade de cada uma destas populações em cada país. Os países A e B possuem Residentes A e Residentes B , respectivamente, com Turistas A/B que se dirigem de A para B e Turistas B/A que se dirigem de B para A . Dessa forma, as influências de origem e destino de turistas foram analisadas através da distância entre os *vetores de mobilidade* dessas populações.

Tabela 1. Comparações entre turistas e residentes de dois países A e B .

Comparação	Comentário
Turista A/B \times Residente A	Como o turista A se comporta em relação à sua origem?
Turista A/B \times Residente B	Como o turista A se comporta em relação ao seu destino?
Turista B/A \times Residente A	Como o turista B se comporta em relação ao seu destino?
Turista B/A \times Residente B	Como o turista B se comporta em relação à sua origem?

Tomando-se um país A como referência, as comparações da Tabela 1 podem ser estendidas para vários países B_i utilizando-se o valor RMS (*Root Mean Square*) das distâncias obtidas pelo método *Canberra*, citadas anteriormente. Por exemplo, tomando-se o Brasil como país A de referência, a primeira comparação da tabela avalia as distâncias $d_i(A/B_i, A)$ entre os vetores de mobilidade dos turistas brasileiros no país B_i e dos residentes no Brasil, e calcula o valor RMS das distâncias d_i , denominado neste trabalho de *Distância Comportamental*. Similarmente, isso é feito para as outras comparações da Tabela 1.

3.6. Identificação de *Returners* e *Explorers*

Para identificação de *Returners* e *Explorers*, cada usuário teve seus *check-ins* agrupados, e foi calculado o raio de giro de todos os locais onde o usuário realizou *check-in*. O raio de giro foi computado de acordo com [Pappalardo et al. 2015], utilizando a equação (1):

$$r_g(k) = \sqrt{\frac{1}{N_k} \sum_{i=1}^k n_i (r_i - r_{cm}^{(k)})^2} \quad (1)$$

sendo $r_g(k)$ o raio de giro das k localidades mais visitadas (limitadas à quantidade total de localidades visitadas), n_i a quantidade de vezes que a localidade i foi visitada, $N_k = \sum_{i=1}^k n_i$, r_i a coordenada bidimensional da localidade i , e $r_{cm}^{(k)}$ o centro de massa, calculado através da média ponderada das coordenadas geográficas de todos os locais

visitados, cujo peso é a quantidade de visitas em cada local. A distância entre os pontos e o centro de massa foi calculada utilizando a fórmula de *Haversine* [Sinnott 1984].

Os usuários foram classificados como *Returners* e *Explorers* através da metodologia apresentada em [Pappalardo et al. 2015]. A classificação consiste em uma fórmula de bisseção, onde o usuário é considerado *Returner* se $r_g(k) > r_g/2$, e *Explorer*, caso contrário, sendo que r_g é o raio de giro de todos os locais visitados. Para este trabalho, utilizou-se $k = 4$.

3.7. *Returners* e *Explorers* em Termos de Categorias de Locais

Com a finalidade de verificar se as mobilidades das classes *Returners* e *Explorers* estão ligadas às preferências de cada usuário, em termos de tipos de locais, foi considerada a categoria estabelecida pelo Foursquare de cada local visitado.

Para cada cidade estudada, os *check-ins* foram separados por nacionalidade do turista. Em seguida cada um desses grupos foi separado em *Returners* e *Explorers*. Logo após, o total de categorias distintas visitadas por cada usuário dessas classes foi contabilizado. Por fim, foi calculada a mediana dessa distribuição de número de categorias distintas por usuários de uma determinada classe.

4. Resultados

Os resultados referentes à mobilidade para efeitos de comparação entre turistas e residentes de cada país são apresentados na Seção 4.1. Os resultados que permitem a análise do comportamento dos usuários medindo a influência regional tendo por base a distância comportamental estão descritos na Seção 4.2. A Seção 4.3 apresenta os resultados referentes à análise de mobilidade no âmbito de *Returners* e *Explorers*.

4.1. Agrupamento de Turistas e Residentes com Relação à Mobilidade

A Figura 1 apresenta o agrupamento através de um dendrograma considerando o seguinte conjunto de países: Brasil (BR), Estados Unidos (US), México (MX), Reino Unido (GB), Turquia (TR), Japão (JP), Indonésia (IN) e Malásia (MY).

Primeiramente, observa-se 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. As exceções observadas são com grupos de japoneses e indonésios.

Um fato interessante é que os residentes da Indonésia parecem ser mais similares aos turistas ocidentais do que aos residentes dos demais países. Além disso, residentes do Japão possuem um comportamento distinto dos demais residentes, mas similar ao de turistas dos países no próprio Japão e de japoneses em outras localidades. Esse fato pode retratar em parte a baixa variabilidade do comportamento de japoneses ao visitarem outros locais em relação ao que fazem no Japão, bem como a força da influência da cultura japonesa em outros países.

Desde já ficam claros, portanto, padrões de turistas e residentes nos agrupamentos que indicam tendências bastante claras de que origem e destino de turistas influenciam sua mobilidade semântica. Esses padrões serão mensurados e melhor analisados na próxima seção.

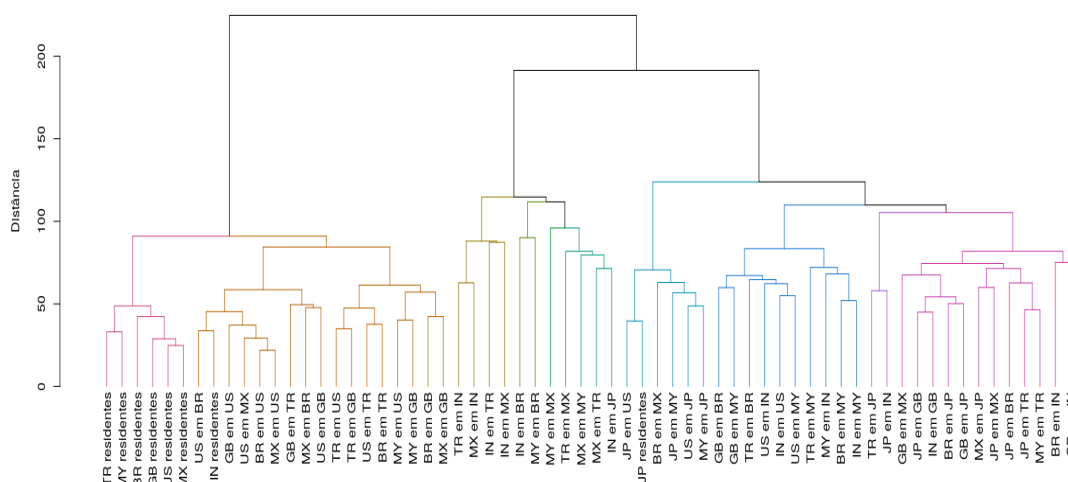


Figura 1. Dendrograma do agrupamento de turistas e residentes nos diversos países.

4.2. Medindo a Influência Regional no Comportamento

Os cálculos das distâncias comportamentais citados na Seção 3.5 foram realizados do ponto de vista de cada país analisado. Primeiramente, são analisadas as comparações “Turista $A/B \times$ Residente A ” e “Turista $A/B \times$ Residente B ”, ou seja, de turistas do país A visitando o país B , comparados com residentes de A e B , respectivamente.

A Figura 2 apresenta esses resultados considerando como origem alguns países e como destino todos os países estudados. Nessa figura os países mencionados no eixo x representam a variável A (ou seja, o país de origem do turista).

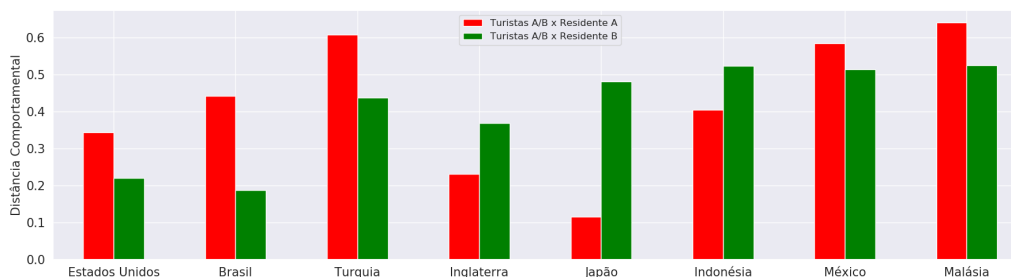


Figura 2. Distância comportamental: turistas A/B (país de origem A indicado).

De acordo com esse resultado, destaca-se o comportamento do Brasil e do Japão. Surge a indicação de que turistas brasileiros se distanciam consideravelmente do comportamento original, ou seja, de residentes no Brasil, passando a absorver de forma mais intensa os costumes locais dos países visitados em contraste com os outros turistas estudados. Já turistas japoneses tendem a manter o seu comportamento original, ou seja, mais próximo do comportamento dos residentes japoneses, ao saírem do país. Esse último resultado também foi parcialmente capturado na análise representada pela Figura 1.

Uma análise similar pode ser feita do ponto de vista de turistas que vêm visitar um determinado país A , utilizando as comparações “Turista $B/A \times$ Residente A ” e “Turista

$B/A \times \text{Residente } B$ ”, ou seja, de turistas do país B visitando o país A , comparados com residentes de A e B , respectivamente.

A Figura 3 mostra esse resultado, onde os países mencionados no eixo x representam a variável A (ou seja, o país de destino do turista). Com essa análise pode-se então ter um melhor entendimento da influência da origem e destino em visitantes de cada país referência.

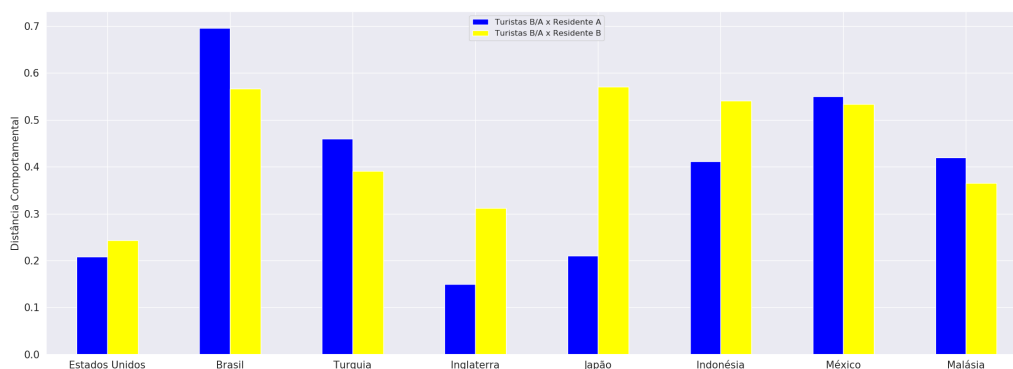


Figura 3. Distância comportamental: turistas B/A (país de destino A indicado).

Lembrando que na Figura 3 os resultados para o Brasil, por exemplo, são referentes a visitantes de todos os países estudados B_i no Brasil comparados com o comportamento dos residentes do Brasil e com o comportamento dos residentes de B_i , ou seja, regras Turistas $B/A \times$ Residentes A e Turistas $B/A \times$ Residente B , respectivamente.

Utilizando novamente os exemplos de Brasil e Japão, percebe-se que visitantes no Brasil demonstram um comportamento menos comparável ao de residentes, tanto de brasileiros, como residentes de suas respectivas origens. Esse efeito pode ser causado por comportamento turístico de lazer, que tipicamente difere daquele de residentes. Enquanto isso, os visitantes no Japão parecem adotar costumes locais. É possível perceber que esses visitantes possuem padrões de mobilidade bastante diferentes daqueles percorridos em seus países de origem. Fortes fatores culturais tradicionais japoneses podem indicar uma causa para este fato.

4.3. Mobilidade de acordo com o Fenômeno de *Returns* e *Explorers*

Na análise apresentada nesta seção duas grandes cidades foram consideradas: New York e Tokyo. A classificação dos usuários em *Returns* e *Explorers* apresentou resultado semelhante àqueles obtidos por Pappalardo et al. (2015). Assim como apresentado por aqueles autores, a Figura 4 mostra que no presente trabalho a quantidade de *Returns* também cresce à medida que *Explorers* diminui, considerando valores crescentes de k .

Essa é uma observação interessante, pois o estudo de Pappalardo e seus co-autores utilizam dados de celulares e sequências de GPS, que tendem a ser mais detalhados do que dados de LBSNs. Isso sugere que o fenômeno capturado pelos autores é mais geral, sendo possível de ser capturado com fontes alternativas, como dados de LBSNs que oferecem maior escalabilidade e são mais fáceis de serem obtidos. Nessa metodologia, a classificação de usuários se concentra inicialmente somente em deslocamentos geográficos.

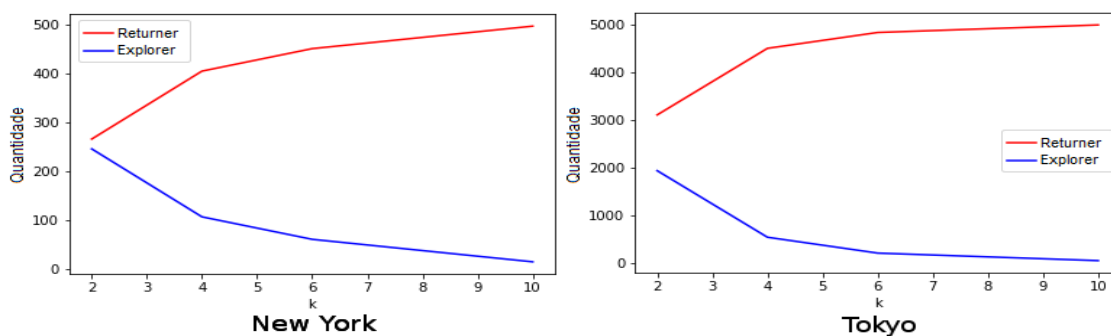


Figura 4. Quantidade de *Explorers* e *Returners* de acordo com k .

Para analisar se a nacionalidade do turista influencia nas classes *Returners* e *Explorers*, os turistas que visitaram New York e Tokyo foram separados nessas duas classes para cada nacionalidade específica. O resultado é apresentado na Figura 5. Nessa figura são apresentados os resultados para usuários da base de todas as nacionalidades que visitaram cada uma dessas cidades. Foram consideradas nacionalidades com mais do que dois representantes distintos. Para New York foram identificados turistas da Alemanha (DE), Argentina (AR), Bélgica (BE), Brasil (BR), Canadá (CA), Chile (CH), Estados Unidos (US), México (MX), Reino Unido (GB) e Turquia (TR). Já para Tokyo, observaram-se turistas dos seguintes países: Brasil (BR), Estados Unidos (US), Filipinas (PH), Tailândia (TH) e Japão (JP).



Figura 5. Percentual de *Returners* e *Explorers* de cada país em New York e Tokyo.

De acordo com essa figura, observa-se que os turistas de todas as nacionalidades tendem a ser mais *Returners* possuindo assim comportamentos mais previsíveis de acordo com a perspectiva analisada (coordenadas geográficas). Tokyo apresentou um resultado semelhante ao de New York, no entanto ainda mais acentuado, ou seja, os turistas nessa cidade tendem a ser ainda mais *Returners*.

Apesar da dominância da classe *Returners*, é possível notar que existem representantes que são *Explorers*. Ter essa informação de antemão pode proporcionar o oferecimento de um serviço de maior qualidade para turistas pertencentes a cada uma dessas classes, por exemplo, em termos de recomendação de roteiros.

Em seguida, foram estendidas as análises apresentadas em [Pappalardo et al. 2015] considerando os tipos de locais (ou categorias, como Restaurante) visitados por essas duas classes de usuários. Foi calculado o valor mediano

referente ao número de categorias distintos visitados por *returners* e *explorers*, segundo a subseção 3.7, para turistas de diversos países em New York e Tokyo. Esse resultado é apresentado na Figura 6, considerando as nacionalidades de turistas estudadas em cada uma dessas cidades.

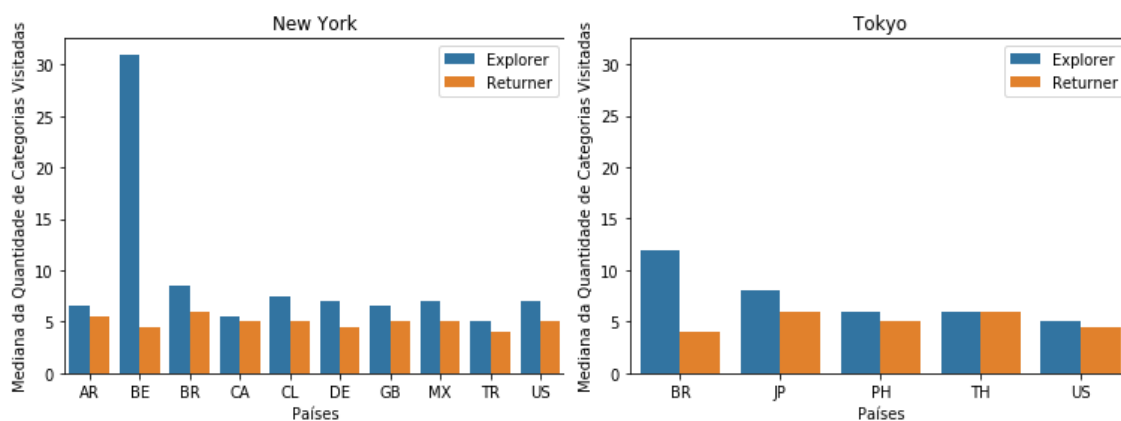


Figura 6. Valor mediano do número de categorias distintas visitadas por *returners* e *explorers* de cada país em New York e Tokyo.

O resultado indica que existe certa disparidade entre a quantidade de categorias visitadas por cada classe, ou seja, os propósitos das visitas são também mais distintos, não somente apresentando localidades diferentes. Isso é válido para a maioria dos países (e em especial para a Bélgica). Com base nos resultados apresentados na Figura 6, pode-se constatar que *Explorers* tendem a visitar uma diversidade maior de categorias, além de visitarem mais localidades físicas distintas, exceto para turistas da Tailândia em Tokyo, onde os valores são idênticos.

É possível ainda observar que esse fenômeno é mais acentuado para turistas de certos países, tanto para New York como Tokyo. Em New York, por exemplo, turistas *Explorers* da Bélgica tendem a visitar muito mais locais de categorias distintas. O mesmo acontece quando se observa Tokyo, onde turistas *Explorers* do Brasil tendem a visitar uma diversidade significativamente maior de categorias. Em Tokyo, turistas *Explorers* dos Estados Unidos e do Brasil não apresentam o mesmo comportamento observado para New York.

Esse resultado sugere uma complementação do modelo proposto por [Pappalardo et al. 2015], uma vez que as mobilidades das diferentes classes também possuem correlação pela preferência de locais. Turistas *Explorers* tendem a visitar, no geral, mais locais distintos enquanto *Returners* visitam os mesmos tipos de locais, e isso também é válido para a semântica dos locais visitados. No entanto, é importante observar que para várias nacionalidades a diferença da diversidade de tipos de locais visitados entre *Returners* e *Explorers* não é muito grande.

Esse resultado indica que é mais fácil prever quais são as categorias que um *Returner* de todos os países irá visitar, e isso é válido para *Explorers* de várias nacionalidades. Esse resultado ainda possui outras implicações. Por exemplo, sabendo a classe de um determinado turista, *Returner* ou *Explorer*, e o seu país de origem, podem ser adaptados sistemas de recomendação que sejam mais condizentes com o que usuários similares

tendem a fazer enquanto turistas em uma determinada cidade.

5. Discussão

Foram analisados resultados agrupados por países para validar a existência de um padrão seguido por cada nacionalidade ao desempenhar papéis diferentes: residentes e turistas. O agrupamento regional entre residentes sugere que existe um comportamento padrão entre países próximos, e a evidente separação de residentes e turistas indica que os padrões mudam de acordo com o local em que o indivíduo está.

Além disso, foi utilizada a classificação sugerida em [Pappalardo et al. 2015] para identificar se existe um padrão recorrente e previsível. Os resultados apresentados na Seção 4, indicam que é possível aplicar a metodologia proposta por [Pappalardo et al. 2015] para classificar usuários, utilizando apenas dados relacionados às suas respectivas mobilidades.

O fato do número de usuários *Returns* ser consideravelmente maior do que de *Explorers*, e de que isso parece se repetir em todos os países analisados levanta a hipótese de que o comportamento dos turistas segue um padrão, sugerindo alto grau de previsibilidade.

Ao analisar esse fenômeno em diferentes nacionalidades, *Returns* representaram acima de 80% dos usuários, em todos os países. No entanto, observa-se algumas diferenças, por exemplo, no caso de brasileiros e americanos. Diferente dos americanos, os brasileiros seguem padrões de mobilidade semelhantes, tanto em New York, como em Tokyo, segundo a visão de *Returns* e *Explorers*.

Um fator que pode influenciar nas escolhas dos turistas, e deve ser estudado no futuro, é a disponibilidade de locais que fazem parte de sua rotina. Um turista indonésio poderá ter dificuldades em encontrar suas preferências culturais em um país do ocidente, devido às diferenças religiosas e preferências gastronômicas, porém, terá mais facilidade em países com um perfil parecido.

Logo, além de estudar a nacionalidade, também é necessário validar qual o impacto do destino do turista. Um caso que pode ilustrar esse efeito é a separação em dois grupos diferentes: turistas em países asiáticos e turistas em países europeus. Um ponto a ser estudado sobre a separação supracitada, é a similaridade com residentes do Japão, que pode indicar que os países asiáticos recebem grande carga de turistas japoneses.

6. Conclusão e Trabalhos Futuros

Utilizando dados de LBSNs, foi possível obter novos resultados relacionados a estudos da mobilidade humana. Para este trabalho, foi considerado um *dataset* extraído da plataforma Foursquare-Swarm.

Considerando-se então as questões de pesquisa levantadas pode-se afirmar que existe similaridade no comportamento de residentes de uma mesma região geográfica. Além disso, turistas tipicamente possuem comportamento influenciado por sua origem, porém bastante distinto daquele dos residentes de seu país, dando forma a padrões de mobilidade próprios. Pode-se ainda ressaltar que a maioria dos turistas possui um padrão de mobilidade previsível. Para entender melhor essa previsibilidade e se ela pode ser utilizada em novos estudos, foi adotada uma métrica para verificar a influência do meio

sobre o comportamento do usuário. Foi observado que influências regionais têm efeitos distintos no comportamento de usuários em determinados países. Os resultados indicam que existe a possibilidade de melhoria de serviços oferecidos a turistas considerando os aspectos estudados.

Para trabalhos futuros, o estudo atual deverá ser estendido, procurando compreender melhor as características de cada país que possuem maior influência na mobilidade de seus turistas e validando a previsibilidade de seus comportamentos a partir de tais características. Tal análise será fundamental em uma futura construção de preditores de rota para turistas, customizada a partir de sua origem e destino. Questões particulares levantadas ao longo da pesquisa também serão exploradas em trabalhos futuros, por exemplo a análise da evolução no comportamento dos usuários urbanos ao longo do tempo e a análise sócio/antropológica para explicar e validar parte dos resultados encontrados.

Agradecimentos

Este trabalho foi parcialmente apoiado pelo projeto CNPq-UrbComp (processo 403260/2016-7), CAPES (código de financiamento 001) e projeto GoodWeb (processo #2018/23011-1 da agência de fomento FAPESP). M. R. Delgado agradece ao CNPq, processos 309935/2017-2 e 439226/2018-0. H. C. M. Senefonte agradece à UEL - Universidade Estadual de Londrina.

Referências

- Falher, G. L., Gionis, A., and Mathioudakis, M. (2015). Where is the soho of rome? measures and algorithms for finding similar neighborhoods in cities. In *Proc. of Int. Conf. on Web and Social Media (ICWSM)*, Oxford, UK.
- Ferreira, A. P. G., Silva, T. H., and Loureiro, A. A. F. (2019). Profiling the mobility of tourists exploring social sensing. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 522–529, Santorini, Greece.
- Gavilanes, R. O. G., Mejova, Y., and Quercia, D. (2014). Twitter ain't without frontiers: economic, social, and cultural boundaries in international communication. In *Proc. of Conf. on Computer-Supported Cooperative Work and Social Computing (CSCW)*, Baltimore, USA.
- Gu, Y., Liu, W., Yao, Y., and Song, J. (2014). Fast routing in location-based social networks leveraging check-in data. In *Proc. of 2014 IEEE Cyber, Physical and Social Computing (CPSCom)*, pages 428–435, Taipei, Taiwan.
- Hristova, D., Williams, M. J., Musolesi, M., Panzarasa, P., and Mascolo, C. (2016). Measuring urban social diversity using interconnected geo-social networks. In *Proc. of Int. World Wide Web Conference*, Montreal, Canada.
- Huang, C. and Wang, D. (2016). Unsupervised interesting places discovery in location-based social sensing. In *Proc. of International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 67–74, Washington, USA.
- Morais, A. and Andrade, N. (2014). The relevance of annotations shared by tourists and residents on a geo-social network during a large-scale touristic event: The case of são joão. In Rossitto, C., Ciolfi, L., Martin, D., and Conein, B., editors, *COOP*

2014 - *Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27-30 May 2014, Nice (France)*, pages 393–408.

- Organization, W. T. (2018). *UNWTO Tourism highlights, 2018 edition*.
- Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., and González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. *EPJ Data Science*, 4(1):1–17.
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature communications*, 6:8166.
- Ribeiro, A. I. J. T., Silva, T. H., de L. P. Duarte-Figueiredo, F., and Loureiro, A. A. F. (2014). Studying traffic conditions by analyzing foursquare and instagram data. In *Proc. of Inter. Symp. on Perf. Eval. of Wireless Ad Hoc, Sensor, and Ubiq. Networks (PE-WASUN)*, Montreal, Canada.
- Rodrigues, D. O., Silva, T. H., Curado, M., Loureiro, A. A., and Villas, L. (2018). Uso de dados de mídias sociais para desenvolvimento de metodologia de posicionamento de táxis. In *Proc. do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Campos do Jordão, Brasil. SBC.
- Santala, V., Miczevski, S., de Brito, S. A., Baldykowski, A. L., Gadda, T., Kozievitch, N., and Silva, T. H. (2017). Making sense of the city: Exploring the use of social media data for urban planning and place branding. In *Proc. do I Workshop de Computação Urbana (COURB)*, volume 1, Belém, Brasil. SBC.
- Santos, F. A., Silva, T. H., Loureiro, A. A., Boukerche, A., and Villas, L. A. (2018). Identificação da reputação de áreas urbanas externas com dados de mídias sociais. In *Proc. do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Campos do Jordão, Brasil. SBC.
- Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. (2017). A large-scale study of cultural differences using urban data about eating and drinking preferences. *Information Systems*, 72(Supplement C):95 – 116.
- 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 Comput. Surv.*, 52(1):17:1–17:39.
- Sinnott, R. W. (1984). Virtues of the haversine. *Sky Telesc.*, 68:159.
- Veiga, D. A., Frizzo, G. B., and Silva, T. H. (2019). Cross-cultural study of tourists mobility using social media. In *Proc. do XXV Simp. Bra. de Sistemas Multimídia e Web (WebMedia)*, pages 313–316, Rio de Janeiro, Brasil.
- Yang, D., Zhang, D., and Qu, B. (2016). Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Trans. Intell. Syst. Technol.*, 7(3):30:1–30:23.
- Zheng, Y., Capra, L., Wolfson, O., and Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.*, 5:38:1–38:55.