Regional Influences on Tourists Mobility Through the Lens of Social Sensing

Helen Senefonte[1,2][0000-0001-8990-9162], Gabriel Frizzo[1][0000-0002-9047-1716], Myriam Delgado[1][0000-0002-2791-174X], Ricardo Lüders[1][0000-0001-6483-4694], Daniel Silver[3], and Thiago Silva[1,3][0000-0001-6994-8076]

[1] Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, Brazil
[2] Universidade Estadual de Londrina, Londrina, Brazil
[3] University of Toronto, Toronto, Canada

Abstract. This study aims at exploring social media data to evaluate how regional and cultural characteristics influence the mobility behavior of tourists and residents. By considering information taken from the mobility graphs of users from different countries, we observe that users' origins can influence their choices. Additionally, the analysis performed in the experiments shows that a regression model could enable the prediction of the behavior of a tourist from a specific country when visiting another country, based on their cultural distances (obtained offline). The ability to explore the cultural characteristics of each nationality in different destinations shows a promising way to improve recommendation systems for points of interest and other services to particular groups of tourists.

Keywords: Cross-Cultural Study · Tourists · LBSNs · Mobility

1 Introduction

Studies based on traditional data sources tend to suffer from poor scalability. Consequently, the experiments are limited, and results are restricted to small regions, like a city or state. The use of location-based social network (LBSN) data can mitigate the scalability problem enabling the study of social behavior for larger populations. These studies contribute to solving problems in many areas ranging from socio-economic issues [2, 3] to assisting tourism-related activities [1, 13].

Planning for urban tourism is important in many aspects, not least economically. Considering data from foreign tourists, world tourism generated revenues of more than one billion US dollars in 2017 alone [6]. Tourism is also a crucial source of jobs in the labor market: according to the World Tourism Organization, it accounts for 1 in 10 jobs [6]. To keep tourist activity in a particular location attractive, it is essential to understand tourists' preferences in order to offer better and smarter services.

With this goal in mind, this research aims at exploring social media data to study regional influences on the behavior of tourists and residents in the context of mobility. The central questions that guide this research are: (i) In terms of mobility, how do tourists behave in countries where they travel? Do they move similarly to residents of destination countries, or do they have similar mobility patterns to residents of their own countries of origin? (ii) Do tourists with the same cultural habits perform similar tourist activities?

To answer the previous questions, we first define the dataset, which encompasses data of eight countries in different regions worldwide, aiming to establish two main categories of users for each country: tourists and residents. The mobility of tourists and residents in distinct countries is analyzed in this study with Foursquare-Swarm check-ins. We formalize the mobility graph to represent residents and tourists' movement at different times throughout the day. Using the type of location visited and the geographic coordinates, this representation captures, in a certain way, the semantic of mobility. After that, we present the concept of *behavioral distance*, which is based on the vectors obtained from the linearization of adjacency matrices representing the mobility graphs of residents and tourists. The behavioral distance allows us to perform two main analyses: (i) the influence of origin and destination on tourists of different countries; (ii) the inference, from a linear regression model, of the behavioral distance of a specific country considering its cultural distance from the other countries.

2 Related Work

Several works on tourist behavior use data obtained from traditional approaches such as questionnaires and interviews. For example, by exploring questionnaire data, Zieba [14] studied how individual features of Austrian tourists can influence their travel motivation. Additionally, the author investigated if tourists' habits are different in other countries. Scuderi and Chiara [8] analyzed expense patterns using a tourist card (which provides discounts in the city of Trentino).

Given the vast amount of data available from LBSN, there is a broad spectrum of possibilities for carrying on studies about urban societies in an unprecedented scale [10], including those focused on cultural aspects. For example, Silva et al. [9] explored Foursquare check-ins on food venues to automatically identify cultural boundaries, whose results are compatible with those obtained using traditional data.

In the context of tourists' behavioral pattern analysis exploring LBSNs data, Vu et al. [12] explore Foursquare check-ins to study the preferences of Malaysian tourists in various countries and their differences from Thai tourists. The study suggests that despite the regional proximity of Malaysia and Thailand, it is possible to find relevant differences in the users' preferences of each country.

Long et al. [5] explore temporal characteristics of tourist check-ins to analyze their movements, finding that most tourists tend to exhibit higher diversity in their activities. Ferreira et al. [1] consider the preferences of tourists and residents represented by Foursquare check-ins and use a spatiotemporal graph model to suggest that each group has distinct preferences in different times for specific places of the studied cities. Veiga et al. [11] extract tourist mobility from two different Foursquare datasets. The authors tested two different ways of identifying tourists and concluded that the results are not significantly influenced by the identification process. In addition, big cities tend to dominate the mobility pattern for the whole country.

The present work differs from the previous ones due to its focus on the analysis of the users' origins influence on their mobility decisions. A comparison between tourists and residents has been carried out, obtaining different groups in terms of mobility patterns. In addition, we quantified regional influence on mobility based on the concept of *behavioral distance*. We found evidence that the user's origin can influence the activities performed. Moreover, we evaluated a model to predict the behavioral distance using as a predictor the cultural dimensions defined by Inglehart and Welzel [4].

3 Data and Method

The considered dataset is publicly available and composed by Foursquare-Swarm check-ins shared on Twitter between January and June 2014 ($\approx$20 million check-ins from $\approx$1.2 million unique users). These data are obtained by a web crawler developed especially for this purpose. Each record of this dataset identifies (by a single ID) the user who performed the activity, the GPS location of the venue visited, the time and category of the venue (names and organization are established by Foursquare-Swarm[4]). Aiming at getting a large data volume and diversity of world regions, the following countries are considered: Brazil (BR), United States (US), Indonesia (ID), Japan (JP), Malaysia (MY), Mexico (MX), the United Kingdom (UK), and Turkey (TR).

It is essential to identify tourists and residents to reach the goals of this study. Our dataset allows us to identify the country of users' residence with minor inconsistencies. We only consider users with a minimum of 2 check-ins, totaling 1010953 unique users. The country of users' residence is chosen as the country where they spent most of the time. Elsewhere, they are considered tourists. In the case of two subsequent *check-ins* are made in different countries, we suppose that the user spends this time window in the first country. Although these premises bring possible misinterpretation, we believe it could have at most a minor impact on the results due to the large volume of data. Previous works in the literature have successfully used similar strategies [7, 1].

Subsequent *check-ins* are modeled in a *mobility graph* which is a directed graph $G_m(V, E)$ obtained for each country m available in our dataset. Vertices $v_i \in V$ represent ten main categories[5] of Foursquare-Swarm. If a user makes two subsequent *check-ins* at v_i and v_j within no more than two periods of the day, a transition or edge $e(i, j) \in E$ connects v_i to v_j. A period of the day includes: (i) morning between 6:00 am and 9:59 am; (ii) noon between 10:00 am and 2:59 pm; (iii) afternoon between 3:00 pm and 6:59 pm; (iv) night between 7:00 pm and 11:59 pm; and (v) dawn between 00:00 am and 05:59 am. For each country, there is one mobility graph for residents and several mobility graphs for tourists of a certain country visiting other countries.

Each mobility graph is represented by an adjacency matrix 10×10 corresponding to ten main categories mentioned above. Each element (i, j) of the matrix corresponds to the number of transitions made by all users between categories i and j. The matrix lines are then concatenated in a 100 position vector called *mobility vector*. Comparisons between mobility vectors are made by *Canberra* distance combined with *Ward* linkage criteria[6].

[4] https://developer.foursquare.com/docs/resources/categories.

[5] Arts & Entertainment; College & University; Professional & Other Places; Residences; Outdoors & Recreation; Shops & Services; Nightlife Spots; Food; Travel & Transport; and Event.

[6] Other distances and linkage criteria were evaluated, but this combination provided the most consistent results.

We are interested in evaluating how tourists are influenced by the countries of their origin and destination. Let's assume two countries O (origin) and D (destination) with mobility vectors defined for Residents O, Residents D, and tourists of O in D denoted by Tourists O/D. We compare these mobility vectors for two cases: (i) Tourists $O/D\times$ Residents O; (ii) Tourists $O/D\times$ Residents D. The first case shows how far (or close) tourists of O are from residents of O when visiting D, whereas the second case shows how tourists of O are far (or close) from residents of D when visiting it. Taking a country O as a reference, mobility vectors for Tourists O/D_i can be built for different destination countries D_i. For example, taking USA as country O, the distance between mobility vectors of USA tourists in country D_i and residents of USA is given by $dist_i(O/D_i, O)$ (similarly for $dist_i(O/D_i, D_i)$ when residents of D_i are considered). The *behavioral distance* (Bh) is then defined in this work as the RMS (*Root Mean Square*) over $dist_i$ calculated with the Canberra distance. Therefore, there are two different behavioral distances: Bh_O and Bh_D depending on whether tourists are compared with their origin O or their destination D_i, respectively.

According to Inglehart and Welzel [4], the *cultural distance* (Ct) is based on the space defined by the two major dimensions of cross-cultural variation: f_{TS} (traditional values versus secular-rational values), and f_{SS} (survival values versus self-expression values). Considering the World Value Survey data from wave 6, 2010-2014, every country is positioned in this two-dimensional space. A particular country O has a cultural distance $Ct(O) = \frac{1}{N}\sum_{i=1}^{N}(euc(O, D_i))$ to a set of countries $\{D_1, D_2, \ldots, D_N\}$, where euc is the euclidean distance. In other words, it is the average Euclidean distance between O and all other countries in the space defined by their cultural dimensions. Finally, we investigate the relationship between cultural and behavioral distances in a set of countries using a linear regression model (ordinary least squares). The behavioral distance (dependent variable) is evaluated from the cultural distance (independent one) by considering two different regression models for Bh_O and Bh_D in the experiments.

4 Results

4.1 Measuring Regional Influence on Behavior

The behavioral distance mentioned in Section 3 is calculated for each analyzed country. Figure 1 presents these results considering the origin of some countries and the destination of all the studied countries. In this figure, the countries mentioned in the x axis represent the variable O (that is, tourist origin country).

According to this result, there is an indication that Brazilian, Mexican, Turkish and Malaysian tourists distance considerably from their original behavior (from residents of their country of origin). Brazilian tourists tend to absorb more intensively the local habits of countries visited in contrast to the other tourists. On the contrary, Japanese tourists tend to maintain their original behavior when leaving the country (close to Japanese residents).

The relative differences between behavioral distances (Bh_O and Bh_D) of Figure 1 is presented in Figure 2. As we can see, Brazil and Japan are the countries with the highest discrepancy, with Brazil being more influenced by the destination and Japan

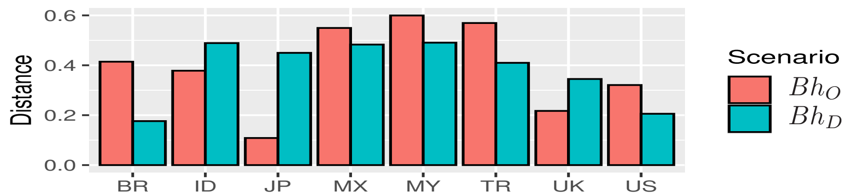


Fig. 1. Behavioral distances Bh_O and Bh_D for tourists of O (BR, ID, JP, MX, MY, TR, UK, US) visiting all other considered countries.

in the opposite way. For the other countries, the discrepancies are lower, especially for Mexico.

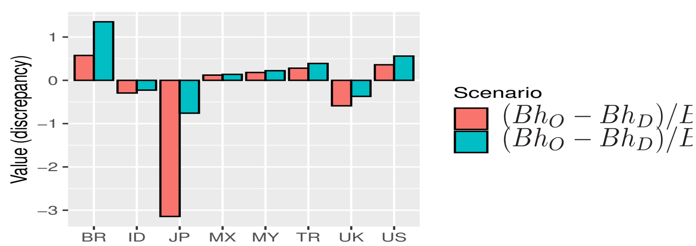


Fig. 2. Relative differences between behavioral distances Bh_O and Bh_D for all countries.

4.2 Regressions on Cultural and Behavioral Dimensions

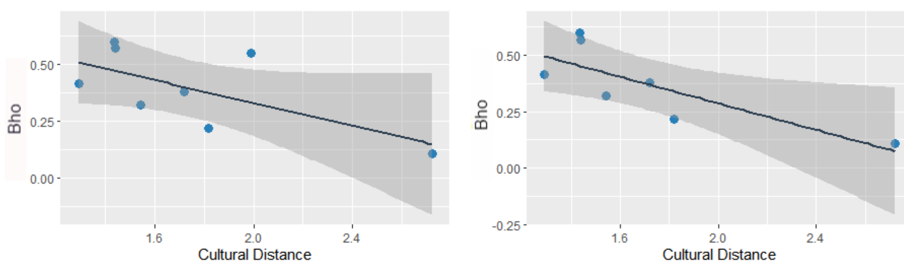
Here we present the results for the linear regression models explained in Section 3. Table 1 summarizes the results. As we can see, the preliminary results suggest that there is a considerable potential of using the cultural distance based on the World Value Survey to explain the behavioral distance, especially in the case Bh_O with $R^2 = 0.42$ (considering all countries) and p-value < 0.1 for all coefficients. This is not the case for Bh_D whose regression models are not relevant with $R^2 = 0.2$ (all countries).

Table 1. Linear regression parameters and quality index R^2, for models predicting the behavioral distances Bh_O and Bh_D for all countries and without Mexico. Significance values codes: ** $p < 0.01$, * $p < 0.05$, $.p < 0.1$.

	All countries		**Without Mexico**	
	Bh_O	Bh_D	Bh_O	Bh_D
(Intercept)	0.832**	0.174	0.876**	0.19
Cultural distance	-0.251.	0.119	-0.294*	0.104
R^2	0.42	0.2	0.64	0.15

As we can see in Figure 3(b), Mexico is an outlier for Bh_O. We tested therefore the linear regression without this country (the results are also presented in Table 1). The message is still the same; however, the quality of the model is better: $R^2 = 0.64$, and the coefficients are significant with 95% of confidence. In the scenario without Mexico, the prediction of Bh_D do not improve.

Moreover, it is interesting to observe the interpretation indicated by models of Bh_O. For people of countries with more distinct cultural habits (bigger cultural distances), the behavioral distance Bh_O tends to decrease i.e., users tend to become more similar to their origins if they have strong cultural differences with locals.



(a) All countries (b) Without Mexico

Fig. 3. Linear regressions for Bh_O considering all countries and removing Mexico as an outlier.

5 Conclusion and Future Work

Using data from Foursquare-Swarm, this study contributes to a better understanding of important tourists' behavioral characteristics. Our results indicate that tourists typically have behavior influenced by their origin (especially when their cultural distance from the locals is large). We also obtained an indication that offline measures of cultural distances, such as those measured by the World Value Surveys [4], could be good predictors for these behavioral differences. This is interesting because those measures are publicly available for many countries. Our results indicate that place recommendation systems for tourists could be leveraged by a tourist behavior estimation which could take into account additional information such as users' origin, destination, and their cultural characteristics.

This study could be expanded in numerous ways. For example, the availability of places that are part of tourists' routine is one factor that can influence their choices and should be studied in the future. An Indonesian tourist may have difficulty finding his cultural preferences in a western country due to religious and gastronomic differences; however, he will find it more comfortable in countries with a similar profile. We are also working to expand the dataset and evaluate other statistical models to better understand the phenomena under study.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This work is also partially supported by the project GoodWeb (Grant #2018/23011-1 from São Paulo Research Foundation - FAPESP), and CNPq, grants 309935/2017-2, 439226/2018-0, and by a Connaught Global Challenge Award.

References

1. Ferreira, A.P.G., Silva, T.H., Loureiro, A.A.F.: Profiling the mobility of tourists exploring social sensing. In: Proc. of DCOSS'19. pp. 522–529. Santorini, Greece (May 2019). https://doi.org/10.1109/DCOSS.2019.00100
2. Hristova, D., Williams, M.J., Musolesi, M., Panzarasa, P., Mascolo, C.: Measuring urban social diversity using interconnected geo-social networks. In: Proc. of WWW'19. Montreal, Canada (2016)
3. Huang, C., Wang, D.: Unsupervised interesting places discovery in location-based social sensing. In: Proc. of DCOSS'16. pp. 67–74. Washington, USA (2016)
4. Inglehart, R., Welzel, C.: Changing mass priorities: The link between modernization and democracy. Perspectives on Politics **8**(2), 551–567 (Jun 2010). https://doi.org/10.1017/s1537592710001258, https://doi.org/10.1017/s1537592710001258
5. Long, X., Jin, L., Joshi, J.: Towards understanding traveler behavior in location-based social networks. In: Proc. of GLOBECOM'13. pp. 3182–3187 (Dec 2013)
6. Organization, W.T.: UNWTO Tourism highlights, 2018 edition (2018). https://doi.org/10.18111/9789284419876
7. Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., González, M.C.: Urban magnetism through the lens of geo-tagged photography. EPJ Data Science **4**(1), 1–17 (2015)
8. Scuderi, R., Dalle Nogare, C.: Mapping tourist consumption behaviour from destination card data: What do sequences of activities reveal? International Journal of Tourism Research **20**(5), 554–565 (2018)
9. Silva, T.H., de Melo, P.O.V., Almeida, J.M., Musolesi, M., Loureiro, A.A.: A large-scale study of cultural differences using urban data about eating and drinking preferences. Information Systems **72**(Supplement C), 95 – 116 (2017). https://doi.org/https://doi.org/10.1016/j.is.2017.10.002
10. Silva, T.H., Viana, A.C., Benevenuto, F., Villas, L., Salles, J., Loureiro, A., Quercia, D.: Urban computing leveraging location-based social network data: A survey. ACM Comput. Surv. **52**(1), 17:1–17:39 (Feb 2019). https://doi.org/10.1145/3301284, http://doi.acm.org/10.1145/3301284
11. Veiga, D.A., Frizzo, G.B., Silva, T.H.: Cross-cultural study of tourists mobility using social media. In: Proc. of WebMedia'19. pp. 313–316. ACM, Rio de Janeiro, Brasil (2019)
12. Vu, H.Q., Li, G., Law, R.: Cross-country analysis of tourist activities based on venue-referenced social media data. Journal of Travel Research p. 0047287518820194 (2019). https://doi.org/10.1177/0047287518820194, https://doi.org/10.1177/0047287518820194
13. Yang, D., Zhang, D., Qu, B.: Participatory cultural mapping based on collective behavior data in location-based social networks. ACM Trans. Intell. Syst. Technol. **7**(3), 30:1–30:23 (Jan 2016). https://doi.org/10.1145/2814575, http://doi.acm.org/10.1145/2814575
14. Zieba, M.: Cultural participation of tourists–evidence from travel habits of austrian residents. Tourism Economics **23**(2), 295–315 (2017)