Using Facebook data to expand multidimensional measures of distance between countries

Extended abstract submitted to the European Population Conference 2020

Carolina C. Vieira
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
carolcoimbra@dcc.ufmg.br

Filipe N. Ribeiro
Universidade Federal de Ouro Preto
Ouro Preto, Brazil
filipe.ribeiro@ufop.edu.br

Pedro O. S. Vaz de Melo
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
olmo@dcc.ufmg.br

Fabrício Benevenuto
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
fabricio@dcc.ufmg.br

Emilio Zagheni
Max Planck Institute for
Demographic Research
Rostock, Germany
zagheni@demogr.mpg.de

ABSTRACT

The gravity law is the inspiration behind one of the most important frameworks used to predict population movement nowadays. One of the most important elements that characterizes the gravity models applied to migration is the distance measure that represents the attraction between regions. The distance considered by this model is not limited to the geographic distance, and can also include administrative and political distance, economic distance, and cultural distance. By using Facebook data, we introduce a methodology to infer cultural elements that immigrants take off to other countries during the migration process in order to measure the distance between countries in terms of its population preferences on its foreign elements. Many cultural aspects characterize regions in terms of cultural attributes, such as clothing, music, art, and food. As an example of application, we focus on cultural elements related to the Brazilian cosine, once the cuisine of a country can effectively reflect one of the dominant aspects of its culture. In this study, we measure the global spread of Brazilian food culture across countries by exploring Facebook user's preferences for typical Brazilian dishes using Facebook data from the Facebook Advertising Platform. The results show a high correlation between the proportion of Brazilian immigrants in each country and the distance between these countries to Brazil in terms of the cultural distance proposed. Because of this, this measure of distance can complement other metrics of distance applied to gravity-type models, for example, in order to explain the flows of people between countries.

1 INTRODUCTION

In social sciences, gravity models are ordinarily applied to studies related to international trade and migration [4, 14, 17]. The most important element in those models is the distance measure that characterizes the attraction between regions. The distance or similarity between two countries can be measured by administrative and political distance, geographic distance, economic distance and also in terms of cultural distance [11]. All of these distance metrics complement other metrics of distance being incorporated in models in order to explain the flows of people between countries. In this work, we focus on developing a measure for cultural distance using social media data provided by Facebook.

There are many cultural aspects that may help us to culturally characterize regions before calculating distance metrics among them, such as preferences for clothes, music, art and food [19]. The cuisine of a country, for example, reflects its history, while the influx of immigrants from many foreign nations develops a rich diversity in food preparation throughout the country[1]. As such, cuisine can be used as a proxy indicator of culture in a country and the number of people interested in a typical national food [1, 2, 20], or even foreign food, could be used to estimate the strength of that culture inside the region.

In this work, Brazilian typical foods are used as a measure of how the Brazilian culture is distributed across various countries in the world. From this, we also measure the cultural distance between Brazil and the countries most preferred by Brazilian immigrants. To do that, we selected 20 typical Brazilian dishes according to a number of websites and collected the data on interests for these dishes from Facebook Advertising Platform (Facebook ads). Using the number of Facebook users interested in certain typical Brazilian food in each country enables us to represent such interests in each country by a vector, what naturally allows us to measure the distance between pairs of countries.

The rest of the paper is organized as follows. In Section 2 we describe the related work. In Section 3 we detail the methodology used to collect the data and describe the data normalization process. The methodology to infer cultural attributes that immigrants take off to other countries is presented in Section 4. This methodology includes a spatial analysis to detect interests that are specific of a country and a comparison of various types of distance between countries considering the selected interests. In Section 5, we present the discussion of our results and offer additional comments about the applicability of the results on Gravity-type models for migration. Finally, we also discuss our findings and draw final conclusions.

[1] https://freelymagazine.com/2017/01/07/what-food-tells-us-about-culture/

2 RELATED WORK

The study of international migration and the development of models to explain and to predict flows of people between countries is not new [17]. And, one of the most important methods is based on the gravity-type models. Cohen et al. [4] developed an algorithm to project future numbers of international migrants from any country or region to any other. Basically, the variables considered by the model include the population and area of origins and destinations of migrants, and the geographic distance between origin and destination. Other researchers point out other distance measures beyond the geographic distance, like administrative and political distance, economic distance and also in terms of cultural distance [11].

In anthropology and sociology, the study of culture can be examined considering a multitude of aspects of our daily life, such as the clothes we wear, the music we listen to, and the food we eat [19]. Food studies are an established interdisciplinary field that recognizes the centrality of food for cultural practices and cultural identity. Considering the dishes from a country or region, for example, it is possible to approximate cultural distance by characterizing the preferences for local foods [1, 2, 20]. Several works, such as [20], explore the idea that food communicates our culture and mechanisms by which we relate food to our cultural identities. Also, Sibal [2] focus on showing the diversity and similarities between people, cultures and food. Similarly, Almerico [1] presented an interdisciplinary study that observes the intricate relationships between food, culture, and society from the sociological perspective.

Once the link between typical food and culture of a country has been established, several works attempted to classify these aspects in order to compare countries in terms of their cultural distance. In [11], the authors list a few types of distances that can be considered when comparing regions and the impacts that each of these distances have, mainly on financial sphere; they found cultural distance to be one of the most important types. Other works, like [5], outline many characteristics that may represent a culture of a country, and used these factors to cluster countries according to their similarity. Silva et al. [21] identify cultural boundaries by analyzing food and drink habits in foursquare. In our work, we focus on Facebook Ads data to measure the similarity between the countries considering only user interests in a small number of typical Brazilian dishes.

The use of the marketing tool provided by Facebook, Facebook Ads, to access and collect data has been recently increasing in many different fields. It was employed in different contexts, such as in predicting crimes [9], musical interests [23], gender gaps [8, 10], migration [6, 22, 24] and to study the relationships between immigrants [13]. But, to the best of our knowledge, this is the first work that uses Facebook data to measure cultural similarity between countries.

3 DATA COLLECTION

Since there is a great variety of typical local foods in Brazil, we selected a set of the 20 most popular Brazilian dishes according to BBC Good Food[2] and the list of Brazilian dishes available on Wikipedia[3]. Due to the fact that our main goal is to compare various countries with Brazil in terms of cultural distance, for this comparison we selected a set containing the 29 countries most preferred by Brazilian immigrants according to the Ministry of Foreign Affairs, Itamaraty[4]. All the subsequent analysis focus on these sets of typical Brazilian dishes and countries.

3.1 Facebook Ads data

The platform to create advertisement on Facebook, Facebook Ads, allows users to compute an estimated audience size for a proposed advertisement [16]. This audience can be defined by demographic attributes provided by Facebook, including gender, age, home location, and interests that can be informed by the user or inferred by Facebook based on user's likes or status updates. Basically, Facebook users can report their preferences over multiple domains such as music, food, and sports [6]. In this paper we focus on typical Brazilian dishes as a marker of cultural distance between countries.

The Facebook Ads API[5] available for Python[6] is used to collect the audience of each food dish on Facebook based on their users' registered preferences. This tool serves perfectly to our work because this audience can also be collected within a given country, as most users have their home location registered in the system. Figure 1 shows the number of users in each location as well as the actual population of the countries. Some countries, like China, will evidently not provide a good estimate for the actual population, given that the number of Chinese users on Facebook is under 0.1% of the real population. Nevertheless, in other countries more than 50% of the real population is part of the Facebook audience.

Before presenting results for food dish audiences, it is important to disclose an API limitation. If the audience size is between 0 (zero) and 1000 (one thousand), the Facebook Ads API will return the default value of 1000. Because of this restriction, the collection considering specific interests inside a small population may not give information about the exact number of Facebook users that match with the criteria specified. Therefore, the comparison in terms of interest in typical Brazilian dishes between Brazilian expats and the rest of the population, especially in countries with small audience on Facebook, is not reliable. Thus, for all cases the API returned the default value, we set the audience to 0.

Figure 1 also shows that the size of audience in each country can vary a great deal across countries. Thus, to make a fair comparison between interests in these countries, we need to normalize the audience in each interest by the estimated Facebook population in each country. More formally, given the Facebook population p of a country c, the audience $A_p(i)$ in c who are interested in dish i is given by:

$$A_p(i) = \frac{audience_p(i)}{p} \tag{1}$$

Figure 2a shows the normalized audience interested in Brazilian dishes for all countries. In this case, note that in all the selected

[2] https://www.bbcgoodfood.com/howto/guide/top-10-foods-try-brazil

[3] https://en.wikipedia.org/wiki/List_of_Brazilian_dishes

[4] http://www.brasileirosnomundo.itamaraty.gov.br/a-comunidade/estimativas-populacionais-das-comunidades/Estimativas%20RCN%202015%20-%20Atualizado.pdf

[5] https://developers.facebook.com/docs/marketing-apis/

[6] https://pypi.org/project/facebookads/

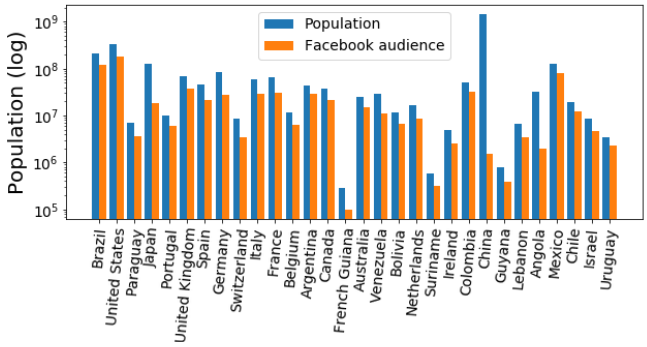


Figure 1: Real population and Facebook audience in each country (log scale).

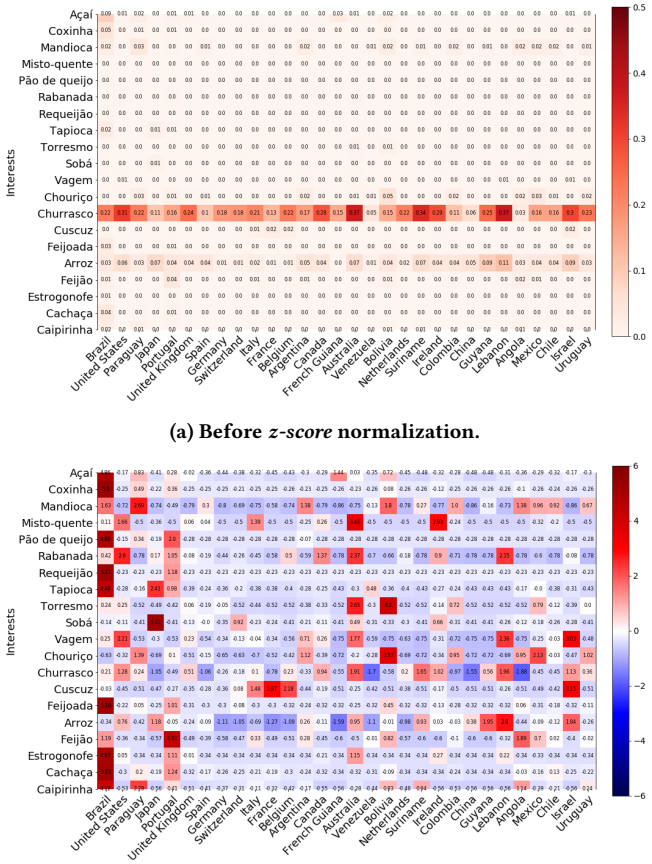
countries, the highest interest is for "Churrasco" ("Barbecue") and, in second, "Arroz" ("Rice"). If we consider each column as a vector representing the country interests, these unbalanced distributions bias the distance measurement between two countries by these two most popular dishes. Note that the difference for the other dishes that have a small proportion of interest in each country is almost zero. In order to give the same importance for all dishes, we normalize and smooth these distributions by their z-scores. Equation 2 shows the formula for calculating the z-scores:

$$z\text{-}score\,(A_p(i)) = \frac{A_p(i) - mean(\mathbf{A(i)})}{std(\mathbf{A(i)})} \tag{2}$$

where $\mathbf{A(i)}$ is the vector that contains $A_p(i)$ for each population p.

Basically, the mean is subtracted from the score for each interest, normalized by the audience in each country, $A_p(i)$, and divided by the standard deviation of the values for that interest in all the countries. As a result, each value now represents the extent to which an interest in a certain country deviates from the mean of a typical distribution. Figure 2b shows the heatmap after *z-score* normalization. As expected, we observe that the distribution is diverse and does not seem to exhibit a few dominant interests in all the countries.

The z-score normalization allows each country to be represented by a vector of preferences regarding typical Brazilian dishes. The aim is to compare those individual vectors with the benchmark Brazilian vector. The most similar countries to Brazil will exhibit small distances. After we generate a set of measures for each country, given by the distance between the country interest vectors and the Brazilian vector, we rank the countries according to the cultural proximity with Brazil. The ranking generated considering the cultural distance can be compared with other types of rankings that attempt to measure the similarity between countries, for instance, the ranking constructed with the most preferred countries by Brazilian immigrants, or Brazilian expats in each country according to Facebook, and rankings that express the geographic distance between countries. The rankings that will be used for this comparison are described next.



(a) Before *z-score* normalization.

(b) After *z-score* normalization.

Figure 2: Proportion of interest in each country. All the interests are normalized by the audience in each country.

3.2 Baseline data

The ranking considering the cultural distance to Brazil can be constructed using different metrics of distance[7]. For each metric, a ranking is generated and compared with the baselines below:

***Immigrant ranking*:** Figure 3a shows the ranking of the countries that have more Brazilian immigrants in proportion to their real populations.

***Expat (Facebook) ranking*:** Figure 3b shows the same countries presented in *Immigrant ranking* sorted by the countries that have more Brazilian expats in proportion to their audience according to Facebook Ads. We can see that both rankings, *Immigrant* and *Expat (Facebook)* are well correlated, while Facebook Ads seems to represent the proportion of Brazilians in those countries well.

***Geographic distance ranking*:** The geographic distance can be expressed in terms of the simple geographic distance or in terms of the weighted distance [18]. The *Geographic distance ranking* is sorted by the countries that are most close to Brazil in terms of a simple geographic distance calculated following the great circle

[7]We decided to use the following distance metrics: Euclidean, Cosine, Mean Absolute error, Relative error and Earth mover's distance

formula, which uses latitudes and longitudes of the most important city in terms of population. *Geographic weighted distance ranking* also shows the countries that are most close to Brazil, considering the distance between the main agglomerations of all countries [12]. Once these two rankings are well correlated (0.96), Figure 3c shows only the *Geographic weighted distance ranking*.

3.3 Comparison between rankings

Table 1 shows correlations between the *Immigrant ranking*, *Expat (Facebook) ranking* and the rankings generated with different measures of distance. Also, the correlations between all of them and the *Geographic weighted distance ranking* are shown in Table 2.

Rankings	WT$^{(i)}$	KT$^{(ii)}$	S$^{(iii)}$	J$^{(iv)}$
Euclidean distance	0.2403	**0.33**	**0.3818**	**0.3333**
Cosine distance	**0.3931**	-0.0739	-0.0995	0.25
Mean Absolute Error	0.3396	0.0788	0.1197	0.1765
Relative Error	0.0099	0.0542	0.103	0.25
Eart Mover's distance	0.1072	0.1823	0.2596	0.25
Expat (Facebook) ranking	0.716	0.0296	0.0655	0.5385
Geo. distance ranking	0.452	-0.0296	-0.0443	0.3333
Geo. W. distance ranking	0.4675	0.0246	0.0128	0.3333

Table 1: Comparison with the *Immigrant ranking*.
(i) WT: Weighted tau correlation.
(ii) KT: Kendall tau correlation.
(iii) S: Spearmanr correlation.
(iv) J: Jaccard similarity considering the top 10 in each ranking.

Rankings	WT$^{(i)}$	KT$^{(ii)}$	S$^{(iii)}$	J$^{(iv)}$
Euclidean distance	0.1896	**0.2611**	**0.3586**	**0.5385**
Cosine distance	0.1833	0.1133	0.2015	0.1765
Mean Absolute Error	**0.2181**	-0.1675	-0.2517	0.25
Relative Error	-0.3877	0.0246	0.0236	0.25
Eart Mover's distance	-0.2908	0.0148	-0.001	0.1111
Expat (Facebook) ranking	0.4563	-0.0099	0.0172	0.4286
Geo. distance ranking	0.9678	0.6601	0.8108	1.0
Immigrant ranking	0.535	0.0246	0.0128	0.3333

Table 2: Comparison with the *Geographic weighted distance ranking*.

In addition to comparing the rankings we are also interested in giving more importance to the first few countries because of their representation in terms of the fraction of Brazilian immigrants in the real population and in the Facebook audience. Hence, we decided to consider the measure of correlation that allots more weight to the top elements in the rank. The weight is mapped from non-negative integers (zero representing the most important element, the first in the *Immigrant ranking*) to a non-negative weight, given by a hyperbolic weighing. The hyperbolic weighting maps the position of each element in rank r to a weight $\frac{1}{r+1}$. Because of this, the first element ($r = 0$) has weight equal to 1, the second, $\frac{1}{2}$, and so on.

As shown in Table 1, considering the `Weighted tau` correlation[8], Cosine distance is the distance metric that better approximates to the *Immigrant ranking* (almost 0.4 correlated). Figure 4 shows the comparison between them. Considering `Kendall tau`[9] and `Spearmanr`[10], the correlation between Euclidean distance ranking and *Immigrant ranking* are higher when compared to other distances. The Cosine distance ranking has a negative correlation when we consider `Kendall tau` and `Spearmanr`. This happens because the last countries in *Immigrant ranking* are associated with a non-matching position in the Cosine ranking list. But when the weights associated to all of the countries are listed in the decreasing order of importance in the *Immigrant ranking*, the mismatches that occur in those positions do not substantially impact the correlation.

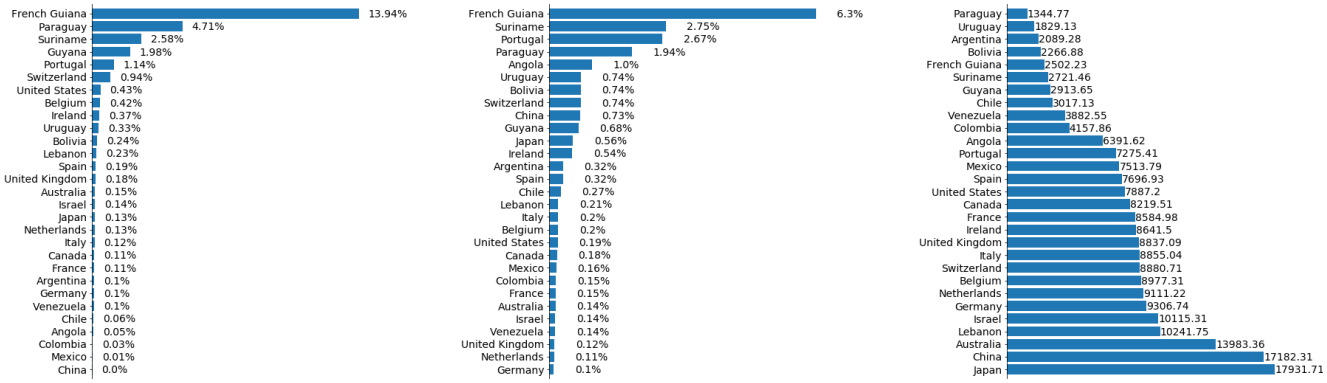
The correlation between the *Immigrant ranking* and the *Expat (Facebook) ranking*, considering the `Weighted tau`, is more than 0.70. This high correlation shows that the Facebook data can be a good estimator of Brazilian immigrants around the world. Also, the correlation between the *Immigrant ranking* and the *Geographic distance ranking* shows that migration are less correlated with the geographic distance, a correlation of 0.45. In fact, it is well know that there are other decisive factors that justify the migration, not only the proximity in terms of geographic distance between countries [7].

Comparing the *Immigrant ranking* and the Cosine distance ranking, shown in Figure 4, we see that despite the large proportion of Brazilian immigrants in countries like Switzerland (6th in *Immigrant ranking*), they seem not to be strongly attached to the Brazilian culture. The opposite is observed in countries like Portugal and Paraguay, that are most preferred by Brazilian immigrants and are most similar in terms of Brazilian food preferences to Brazil. Portugal is the country most similar to Brazil in terms of the preferences for the typical Brazilian dishes. This similarity cannot be related to the geographic distance, since Portugal is not close to Brazil, but the proportion of immigrants in Portugal is one of the highest according to the *Immigrant Ranking*. The language and the general cultural similarity [15] shared between the former colony of Portugal, Brazil, explain in part the pull factors of migration to Portugal. Considering the United States, the country with more Brazilian immigrants in terms of absolute value, the position in the *Immigrant ranking* and in the Cosine distance ranking remains the same, so the similarity in terms of Brazilian food preferences are well correlated with the number of Brazilian immigrants in the population. Other countries, like Paraguay and Bolivia, seem to be more similar to Brazil in terms of food interests because of the geographic proximity. In general, most of the countries like Argentina, Venezuela, Colombia and Chile, that are geographically closer to Brazil are at a higher position relative to the cultural distance ranking, as compared to their *Immigrant ranking*. In this case, this result shows that the interests can be justified not only by the migration processes but also by the geographic proximity, given that the local population seems to be interested in several typical dishes from the neighboring countries. However, generally the migration process is one of the most crucial factors that exposes distant countries (from Brazil) to the Brazilian cuisine.

[8] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.weightedtau.html

[9] docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.stats.kendalltau.html

[10] docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.spearmanr.html



(a) ***Immigrant ranking*****: Proportion of Brazilian immigrants in real population.**

(b) ***Expat (Facebook) ranking*****: Proportion of Brazilian expats in Facebook audience.**

(c) ***Geographic weighted distance ranking*****: Weighted distance to Brazil.**

Figure 3: ***Immigrant, Expat (Facebook) rankings*** **and** ***Geographic weighted distance ranking.***

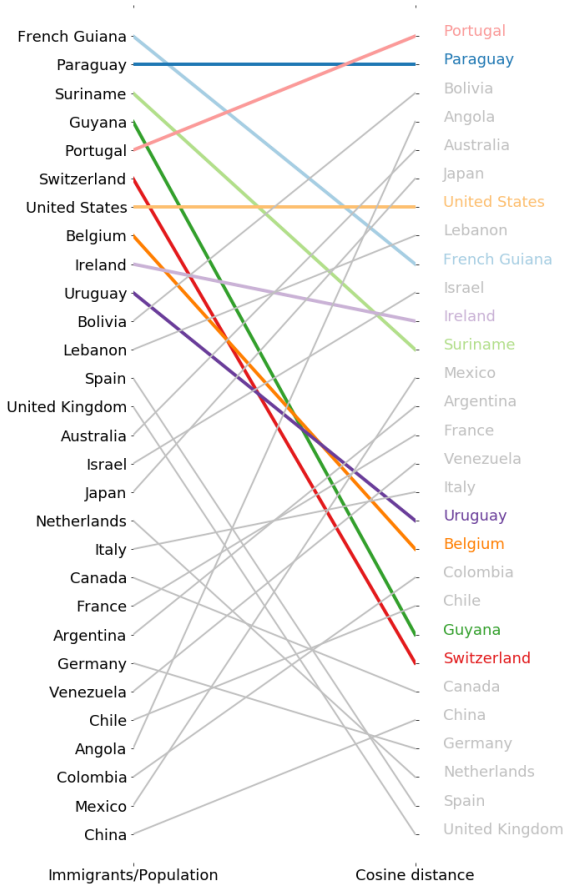


Figure 4: Comparison between ***Immigrant*** **and Cosine distance rankings.**

4 SPATIAL ANALYSIS

Figure 2b shows the proportion of Facebook users interested in some typical Brazilian dishes after the *z-score* normalization. In general, for most of the interests, Brazil is above average. "Chouriço", "Arroz" ("Rice", in English), "Sobá" and "Cuscuz" are the only interests below average, indicating that these interests may be more popular in other countries. Considering only the interests that have the highest value for Brazil, 9 dishes[11] are considered typical from Brazil. In this section, this methodology will be compared to an information theoretic approach that considers spatial analysis [3] to infer typical cultural elements.

Interest entropy and *Interest focus* give us an idea of how the interests are distributed around the globe. Given that $A_p(i)$ is the audience of a region p (in a total of M regions) interested in i, the metric *interest focus* is given by the Equation 3. This measure describes the proportion of the audience that are interested in i in a specific location. On the other hand, Equation 4 shows the *interest entropy* formula that corresponds to the entropy measure of an interest distributed over the countries.

$$F_i = \frac{A_p(i)}{\sum_{p=1}^{M} A_p(i)} \tag{3}$$

$$H_i = -\sum_{p=1}^{M} F_i log_2 F_i \tag{4}$$

These metrics provide support for evaluating the spread of various interests around the globe. Interests that are very concentrated in a region tends to manifest a low *interest entropy* and a high *interest focus* considering that region. It is particularly enticing to look at the interests that have moderate *interest entropy* because these interests are not entirely local, while also not completely common around the world. These interests could originate from a specific region and spread in popularity across other regions that share different cultural aspects, or become popular due to migration.

[11] "Açaí", "Coxinha", "Pão de Queijo", "Requeijão", "Tapioca", "Feijoada", "Estrogonofe", "Cachaça" and "Caipirinha"

For each interest, the difference between the *interest entropy* considering and not considering Brazil in the vector of countries expresses changes related to the uncertainty of which country is the dish associated with. When we consider Brazil, the entropy tends to be lower if the interest is more popular in Brazil than in other countries. Figure 5 shows the result when the entropy difference is calculated and for which dishes the *interest entropy* decrease. Notice that according to this methodology, 7 dishes[12] are considered typical from Brazil, and all of them are also considered by the first methodology using the *z-score* normalization matrix.

Finally, we compare the Brazilian typical interests according to these approaches with the 6 Brazilian typical dishes listed by BBC Good Food. Considering the *z-score* normalization matrix and the *interest entropy* difference, we identify 5 and 4 common interests, respectively. The other 2 dishes from BBC Good Food that are not consider by our metrics as typical from Brazil, "Churrasco" and "Pão de Queijo", seems to be popular in other countries as shown by the metric *interest focus* in Figure 6. We see that "Churrasco" has a uniform distribution over the countries, and in fact, it is not only popular in Brazil. Analyzing "Pão de Queijo", only Brazil and Portugal demonstrate a significant interest for this food, because of this, the uncertainty increase when we do consider Brazil. This result shows that the *interest entropy* difference does not depend only of the highest *z-score* but also considers the whole *interest focus* distribution.

Due to the similarity between the results generated with both methodologies, only the results considering the Brazilian typical dishes, according to the *interest entropy* difference, are reported. Table 3 shows correlations between the *Immigrant ranking*, *Expat (Facebook) ranking* and the rankings generated with different measures of distance considering only those 7 interests, and table 4 shows the correlation between all of them and the *Geographic weighted distance ranking*. When we compare those rankings considering the `Weighted tau` correlation, Cosine distance ranking has a 0.40 correlation with the *Immigrant ranking*. Notice that while our methodology allows us to reduce the vectors size from 20 to 7 dishes, the correlations between the rankings is kept the same.

Rankings	WT$^{(i)}$	KT$^{(ii)}$	S$^{(iii)}$	J$^{(iv)}$
Euclidean distance	0.3786	0.0493	0.0562	0.3333
Cosine distance	**0.3897**	-0.0985	-0.1374	0.3333
Mean Absolute Error	0.3881	-0.0837	-0.1453	**0.4286**
Relative Error	0.3861	**0.0542**	**0.0596**	0.3333
Eart Mover's distance	0.3881	-0.0837	-0.1453	**0.4286**

Table 3: Comparison with the *Immigrant ranking*. Considering only the 7 interests typical from Brazil according to Entropy difference.

5 DISCUSSION

In the literature, many measures of distance such as the geographic distance, typically included in gravity-type models of migration, were found useful to characterize similarities between countries. But, not only does the geographic distance affect migration or trade between countries, administrative and political distances, as well as economic and cultural distances all correspond to other ways to measure similarity. The focus of the present work is to explore specific cultural attributes in order to develop a measure of distance that would most accurately characterize cultural affinities between countries with regard to food preferences. By using social media data to characterize the interests of each country enables us to represent it in terms of its cultural composition and to compare the countries by calculating the aforementioned types of distance between them. Such approach complements the previous research that employs various measures of distance in order to explain the international migration. Finally, the methodology we adopted in the present paper helps one to understand the study of cultural attraction from the social media perspective and the cultural distance between countries can also be included as one more attribute in classic gravity-type model of migration [4].

[12] "Açaí", "Coxinha", "Tapioca", "Feijoada", "Estrogonofe", "Cachaça" and "Caipirinha"

Rankings	WT$^{(i)}$	KT$^{(ii)}$	S$^{(iii)}$	J$^{(iv)}$
Euclidean distance	**0.1998**	**0.1872**	**0.266**	0.3333
Cosine distance	0.1688	-0.0394	-0.03	0.3333
Mean Absolute Error	0.1755	-0.0049	-0.0074	0.3333
Relative Error	0.1763	-0.0148	0.003	**0.4286**
Earth Mover's distance	0.1755	-0.0049	-0.0074	0.3333

Table 4: Comparison with the *Geographic weighted distance ranking*. Considering only the 7 interests typical from Brazil according to Entropy difference.

The Facebook data is a reliable proxy to study international migration. In this study, the correlation between the proportion of Facebook users that are Brazilian expats living abroad and the official data about the number of Brazilian immigrants is more than 0.7. Also, this data is good to estimate interests related to foreign culture, in this case, the Brazilian cuisine. The ranking generated considering the similarity between countries given by food interest is almost 0.4 correlated with the ranking generated with the official proportion of Brazilian immigrants in some countries. We can conclude that the cultural similarity between Brazil and some countries occur due to aspects like the geographic proximity (e.g. Paraguay, Argentina), linguistic similarity (e.g. Portugal, Angola), and most important, due to the number of immigrants in the population, which increases the spread of cultural elements from the country of origin.

REFERENCES

[1] Gina M Almerico. 2014. Food and identity: Food studies, cultural, and personal identity. *Journal of International Business and Cultural Studies* 8 (2014), 1.

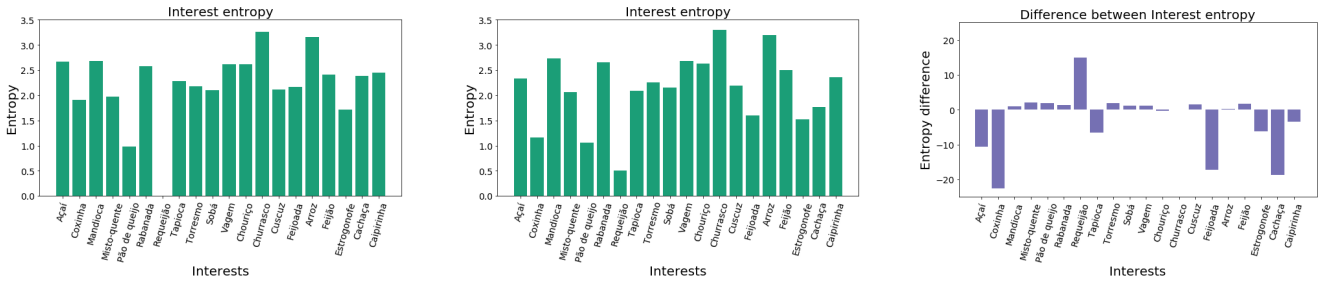
[2] Jean Boutaud, Anda Becuț, and Angelica Marinescu. 2016. Food and culture. Cultural patterns and practices related to food in everyday life. *International Review of Social Research* 6 (2016), 1–3. https://doi.org/10.1515/irsr-2016-0001

[3] Anders Brodersen, Salvatore Scellato, and Mirjam Wattenhofer. 2012. YouTube Around the World: Geographic Popularity of Videos. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*. ACM, New York, NY, USA, 241–250. https://doi.org/10.1145/2187836.2187870

[4] Joel E. Cohen, Marta Roig, Daniel C. Reuman, and Cai GoGwilt. 2008. International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences* 105, 40 (2008), 15269–15274. https://doi.org/10.1073/pnas.0808185105 arXiv:https://www.pnas.org/content/105/40/15269.full.pdf

[5] Gustavo De Santis, Mauro Maltagliati, and Silvana Salvini. 2015. A Measure of the Cultural Distance Between Countries. *Social Indicators Research* 126 (03 2015). https://doi.org/10.1007/s11205-015-0932-7

[6] Antoine Dubois, Emilio Zagheni, Kiran Garimella, and Ingmar Weber. 2018. Studying Migrant Assimilation Through Facebook Interests. In *Social Informatics*,



(a) Interest entropy. **(b) Interest entropy considering Brazil.** **(c) Difference between interest entropy.**

Figure 5: Comparison between interest entropy not considering and considering Brazil.

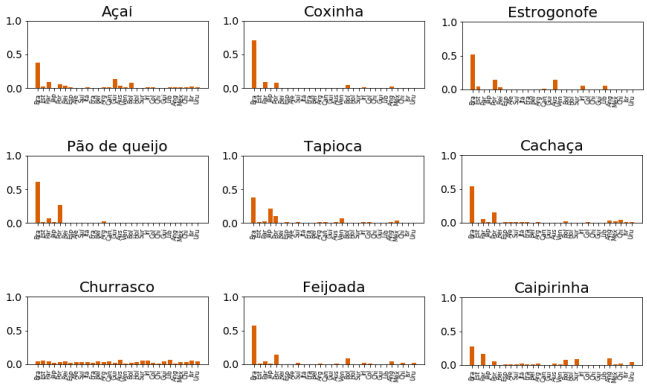


Figure 6: Interest focus considering Brazil.

Steffen Staab, Olessia Koltsova, and Dmitry I. Ignatov (Eds.). Springer International Publishing, Cham, 51–60.

[7] Thomas Faist. 2000. The volume and dynamics of international migration and transnational social spaces. (2000).

[8] Masoomali Fatehkia, Ridhi Kashyap, and Ingmar Weber. 2018. Using Facebook ad data to track the global digital gender gap. *World Development* 107 (2018), 189 – 209. https://doi.org/10.1016/j.worlddev.2018.03.007

[9] Masoomali Fatehkia, Dan O'Brien, and Ingmar Weber. 2019. Correlated impulses: Using Facebook interests to improve predictions of crime rates in urban areas. *PLOS ONE* 14, 2 (02 2019), 1–16. https://doi.org/10.1371/journal.pone.0211350

[10] David Garcia, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. 2018. Analyzing gender inequality through large-scale Facebook advertising data. *Proceedings of the National Academy of Sciences* 115, 27 (2018), 6958–6963. https://doi.org/10.1073/pnas.1717781115 arXiv:https://www.pnas.org/content/115/27/6958.full.pdf

[11] Pankaj Ghemawat. 2001. Distance still matters. *Harvard business review* 79, 8 (2001), 137–147.

[12] Keith Head and Thierry Mayer. 2002. *Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade.* Vol. 1. Citeseer.

[13] Amaç Herdağdelen, Bogdan State, Lada Adamic, and Winter Mason. 2016. The Social Ties of Immigrant Communities in the United States. In *Proceedings of the 8th ACM Conference on Web Science (WebSci '16).* ACM, New York, NY, USA, 78–84. https://doi.org/10.1145/2908131.2908163

[14] Walter Isard. 1954. Location theory and trade theory: short-run analysis. *The Quarterly Journal of Economics* (1954), 305–320.

[15] Bruce Kogut and Harbir Singh. 1988. The Effect of National Culture on the Choice of Entry Mode. *Journal of International Business Studies* 19, 3 (01 Sep 1988), 411–432. https://doi.org/10.1057/palgrave.jibs.8490394

[16] Michal Kosinski, Sandra C Matz, Samuel D Gosling, Vesselin Popov, and David Stillwell. 2015. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist* 70, 6 (2015), 543.

[17] Douglas S. Massey, Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor. 1993. Theories of International Migration: A Review and Appraisal. *Population and Development Review* 19, 3 (1993), 431–466. http://www.jstor.org/stable/2938462

[18] Thierry Mayer and Soledad Zignago. 2011. Notes on CEPII's distances measures: The GeoDist database. (2011).

[19] Ettore Recchi and Adrian Favell. 2019. *Everyday Europe: Social transnationalism in an unsettled continent.* Policy press.

[20] Vatika Sibal. 2018. Food: Identity of culture and religion. 6 (2018), 10908–10915.

[21] Thiago H Silva, Pedro OS Vaz de Melo, Jussara M Almeida, Mirco Musolesi, and Antonio AF Loureiro. 2014. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *Eighth International AAAI Conference on Weblogs and Social Media.*

[22] Spyridon Spyratos, Michele Vespe, Fabrizio Natale, Ingmar Weber, Emilio Zagheni, and Marzia Rango. 2018. Migration data using social media: a European perspective. *Publications Office of the European Union* (2018).

[23] Ian Stewart, René D. Flores, Timothy Riffe, Ingmar Weber, and Emilio Zagheni. 2019. Rock, Rap, or Reggaeton?: Assessing Mexican Immigrants' Cultural Assimilation Using Facebook Data,. In *The World Wide Web Conference (WWW '19).* ACM, New York, NY, USA, 3258–3264. https://doi.org/10.1145/3308558.3313409

[24] Emilio Zagheni, Ingmar Weber, Krishna Gummadi, et al. 2017. Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review* 43, 4 (2017), 721–734.