Using Facebook Data to Measure Cultural Distance between Countries: the Case of Brazilian Cuisine

Carolina C. Vieira
UFMG
carolcoimbra@dcc.ufmg.br

Filipe N. Ribeiro
UFOP
filipe.ribeiro@ufop.edu.br

Pedro O. S. Vaz de Melo
UFMG
olmo@dcc.ufmg.br

Fabrício Benevenuto
UFMG
fabricio@dcc.ufmg.br

Emilio Zagheni
MPIDR
zagheni@demogr.mpg.de

ABSTRACT

Measuring the affinity to a particular culture has been an active area of research. Countries and their residents can be characterized by many cultural aspects, such as clothing, music, art and food. As one of the central aspects, the cuisine of a country can reflect one of the dominant aspects of its culture. As such, the number of people interested in a typical national dish can be used to estimate the prevalence of that culture inside the host region. In this study, we measure the global spread of Brazilian culture across countries by exploring Facebook user's preferences for typical Brazilian dishes through the Facebook Advertising Platform. To decide which dish will be considered typical from Brazil, we made use of spatial analysis to understand the distribution of interests around the world and to quantify how typical the dish is in Brazil and among Brazilian immigrants. This methodology can be generalized to other countries to infer cultural elements that emigrants usually take to and preserve in the countries they migrate to. Also, the interest in Brazilian typical dishes can be used to characterize countries in terms of Brazilian cultural exposition. While evaluating the cultural distance between Brazil and the countries with more Brazilian immigrants, we explore several measures of distance to compare these in the context of affinity to Brazilian cuisine. Our results revealed that these cultural distance measures can complement other metrics of distance applied to gravity-type models, for example, in order to explain flows of people between countries.

CCS CONCEPTS

• **Information systems** → **Web and social media search**; *Social networks*; • **Networks** → **Online social networks**; • **Applied computing** → *Sociology*.

KEYWORDS

social media, social networks, sociology

ACM Reference Format:
Carolina C. Vieira, Filipe N. Ribeiro, Pedro O. S. Vaz de Melo, Fabrício Benevenuto, and Emilio Zagheni. 2020. Using Facebook Data to Measure Cultural Distance between Countries: the Case of Brazilian Cuisine. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3366423.3380082

WWW '20, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380082

1 INTRODUCTION

In social sciences, gravity models are ordinarily applied to studies related to international trade and migration [5, 15, 18]. 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, geographical distance, economic distance and also in terms of cultural distance [12]. There are many aspects that may help us to culturally characterize regions before calculating distance metrics among them, such as preferences for clothes, music, art and food [20].

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 for culture in a country and the number of people interested in a typical national food [1, 3, 23], or even foreign food, could be used to estimate the strength of that culture inside the region.

In this work, Brazilian typical dishes are used as a measure of how the Brazilian culture is distributed around the world. From this, we also measure the cultural distance between Brazil and the countries with more Brazilian immigrants. To do that, we selected 20 typical Brazilian dishes according to Web sources like Wikipedia and collected the data on interests for these dishes through the 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 via a vector. This allows us to evaluate, in a natural way, the distance between pairs of countries. This measure of distance is a new way to express the attraction between countries in terms of cultural distance and it can be incorporated into gravity-type models.

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

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

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.

2 RELATED WORK

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 [20]. 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, 3, 23]. Several papers, such as [23], explore the idea that food communicates our culture and mechanisms by which we relate food to our cultural identities, while others revealed that people and societies can be discriminated by their food and cultural habits [1, 3].

Once the link between typical food and culture of a country has been established, several papers attempted to classify these aspects in order to compare countries in terms of their cultural distance. In [12], 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 the financial sphere: they found cultural distance to be one of the most important factors. Another line of literature [6, 24] outlines many characteristics that may represent a culture of a country and used these factors to cluster countries according to their cultural similarity. In our work, we focus on Facebook Ads data to measure the similarity between countries considering only user interests for a small number of typical Brazilian dishes. To the best of our knowledge, this is the first analysis that uses Facebook data to measure the cultural similarity between countries.

The use of the marketing tool provided by Facebook, namely 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 [10], musical interests [26], gender gaps [9, 11], migration [7, 25, 27], political science [21, 22], and to study the relationships between immigrant communities [14]. Similar to us, Stewart *et al.* [26] discuss the cultural diffusion or the cultural assimilation of immigrants. However, our work compares countries at a high level by analyzing the population interests in each country as a whole. We do not analyze separately the interests of immigrants and natives. Nevertheless, we are aware that this analysis is important for future research in terms of quantifying the extent to which immigrants affect the culture of natives.

3 DATA COLLECTION

Since there is a great variety of typical local food 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 with most Brazilian immigrants according to the Ministry of Foreign Affairs, Itamaraty[4]. All the subsequent analyses focus on these sets of typical Brazilian dishes and countries.

3.1 Facebook Ads data

The platform to create advertisement on Facebook, namely Facebook Ads, allows users to compute an estimated audience size for a proposed advertisement [17]. This audience can be defined by demographic attributes provided by Facebook, including gender, age, home location, and interests, which can be informed by the user or inferred by Facebook based on users' likes or status updates. Basically, Facebook users generate traces of their preferences over multiple domains such as music, food, and, sports [7]. 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 dish on Facebook based on users' preferences. This tool serves our purposes well 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 culinary dish audiences, it is important to disclose the API limitations as well as the steps taken to preserve the privacy of users. First of all, it is important to point out that, among the millions of requests we performed to the Marketing API, none of them were able to gather and link any personal information to any particular user. Our work uses only aggregate information and, if the audience size is between 0 (zero) and 1,000 (one thousand), the Facebook Ads API will return the default value of 1,000. Because of this restriction, data collection considering specific interests for a small population (less than 1,000 users) may not give information about the exact number of Facebook users that match 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 audiences on Facebook, may be too noisy. Thus, for all cases where the API returned the default lower bound value, we set the audience to 0. In addition to this, the mechanisms behind the tool to infer demographic attributes from the offline world are not publicly known, which is a limitation of our method as we cannot evaluate the extent to which the data are reliable. Finally, the population of Facebook is known to be biased with respect to gender, age, and other socio-demographic characteristics as previously discussed by some authors [2].

Figure 1 shows that the audience prevalence can vary a great deal across countries. Thus, to make a fair comparison between

[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/

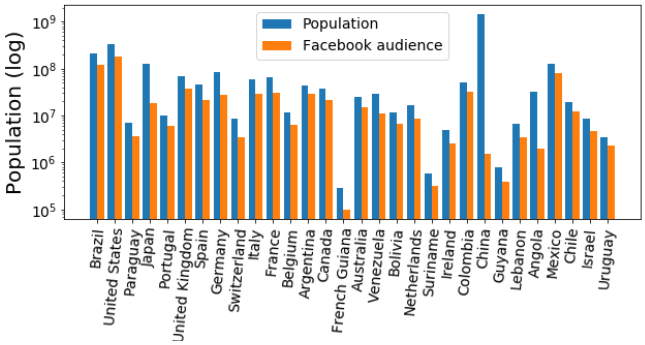


Figure 1: Real population and Facebook audience in each country (log scale).

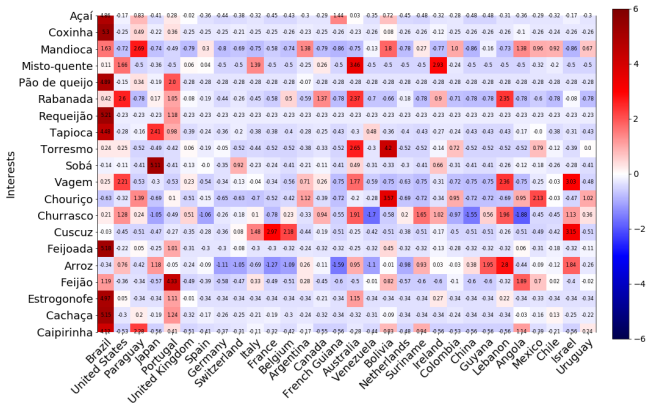


Figure 2: Proportion of interest in each country after *z-score* normalization. All the interests are normalized by the audience in each country.

interests in these countries, we need to normalize the audience in each interest by the estimated Facebook population of 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:

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

Considering the normalized audience interested in Brazilian dishes, for all 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. 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 to all dishes, we normalize and smooth these distributions by their 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 of each interest and divided by its standard deviation. As a result, each value now represents how many standards deviations an interest in a certain country deviates positively or negatively from the mean. Figure 2 shows the heatmap after *z-score* normalization. As expected, we observe that the distribution is heterogeneous 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 of 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 of countries with more Brazilian immigrants and expats according to Facebook, and rankings that express the geographic distance between countries.

3.2 Baseline data

The ranking considering the cultural distance to Brazil can be constructed using different distance metrics, and in this work we use: Euclidean, Cosine, Mean Absolute error, Relative error and Earth mover's distance. 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.

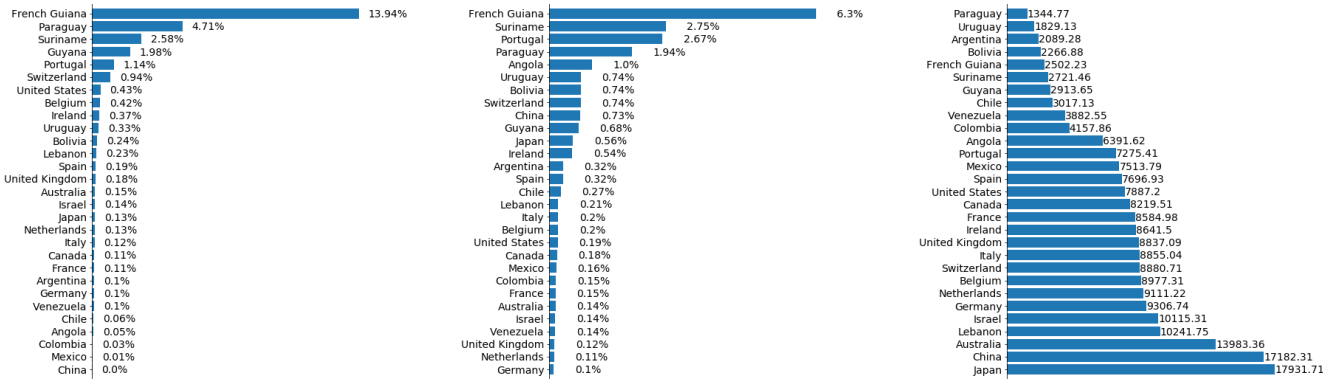
***Expat ranking*:** Figure 3b shows the same countries presented in *Immigrant ranking* sorted by the countries that have more Brazilian expats in proportion to their audience according to Facebook Ads. We can see that both rankings, *Immigrant* and *Expat* are well correlated, while Facebook Ads seems 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 [19]. 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 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 closest to Brazil, considering the distance between the main agglomerations of all countries [13]. 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 and p-values between the *Immigrant ranking*, *Expat 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.

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



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

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

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

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

Rankings	WT(i)	KT(ii)	S(iii)	J(iv)
Euclidean dist.	0.24	**0.33** (0.01)	**0.38** (0.04)	**0.33**
Cosine dist.	**0.39**	-0.07 (0.59)	-0.1 (0.61)	0.25
Mean Abs. Error	0.34	0.08 (0.56)	0.12 (0.54)	0.18
Relative Error	0.01	0.05 (0.7)	0.10 (0.6)	0.25
Eart Mover's dist.	0.11	0.18 (0.17)	0.26 (0.17)	0.25
Expat ranking	0.7	0.03 (0.84)	0.07 (0.74)	0.54
Geo. dist. ranking	0.45	-0.03 (0.83)	-0.04 (0.82)	0.33
Geo. W. dist. ranking	0.47	0.02 (0.87)	0.01 (0.95)	0.33

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 dist.	0.19	**0.26** (0.05)	**0.36** (0.06)	**0.54**
Cosine dist.	0.18	0.11 (0.4)	0.20 (0.3)	0.18
Mean Abs. Error	**0.22**	-0.17 (0.21)	-0.25 (0.19)	0.25
Relative Error	-0.39	0.02 (0.87)	0.02 (0.9)	0.25
Eart Mover's dist.	-0.29	0.01 (0.93)	0.0 (0.99)	0.11
Expat ranking	0.46	-0.01 (0.95)	0.02 (0.93)	0.43
Geo. dist. ranking	0.99	0.66 (0.0)	0.81 (0.0)	1.0
Immigrant ranking	0.53	0.02 (0.87)	0.01 (0.95)	0.33

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

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 that 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[7], Cosine distance is the distance metric that generates the Cultural distance ranking that best approximates the *Immigrant ranking* (almost 0.4 correlated). Figure 4 shows the comparison between them. Considering `Kendall tau`[8] and `Spearmanr`[9], 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 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 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 explain migration, not only the proximity in terms of geographic distance between countries [8].

Figure 4 shows the correlation between migration and cultural distance. The *Immigrant ranking* corresponds to the percentage of Brazilian immigrants in some countries and the cultural distance is given by the cosine distance between countries to Brazil. The distance is calculated between vectors generated for each country considering the percentage of Facebook users who are interested in Brazilian typical food. The most similar countries to Brazil are at the top of the second list in Figure 4. This figure should be interpreted as a comparison between rankings where the top 10 countries in the *Immigrant ranking* are colored to make the comparison easier.

Comparing the *Immigrant ranking* and the Cultural 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

[7] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.weightedtau.html
[8] docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.stats.kendalltau.html
[9] docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.spearmanr.html

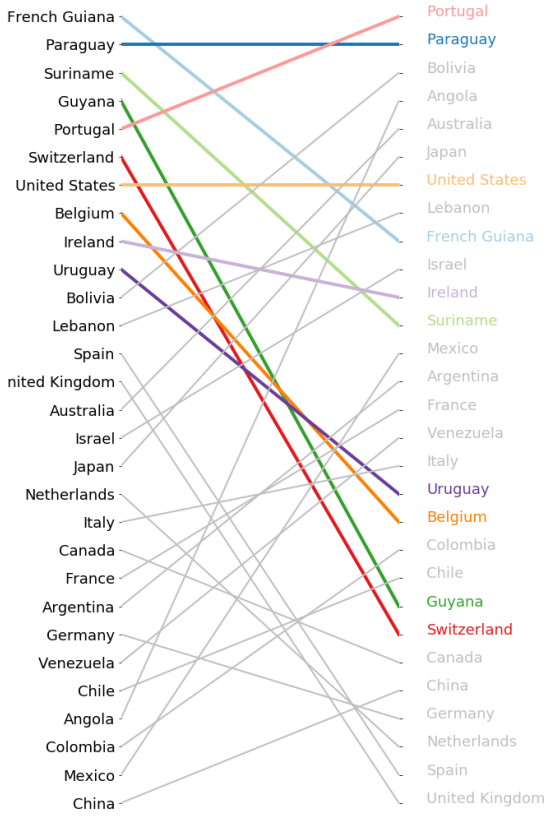


Figure 4: Comparison between *Immigrant* and Cultural distance rankings.

highest according to the *Immigrant Ranking*. The language and the general cultural similarity [16] 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 Cultural 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, which 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.

4 SPATIAL ANALYSIS

Figure 2 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. If we consider that typical Brazilian foods are those in which the interest z-scores in Brazil are the highest among other countries, then 9 dishes meet this criterion: "Açaí", "Coxinha", "Pão de Queijo", "Requeijão", "Tapioca", "Feijoada", "Estrogonofe", "Cachaça" and "Caipirinha". In this section, this approach will be compared to an information theoretic approach that considers spatial analysis [4] to infer typical cultural elements.

This approach is based on two metrics, namely *Interest entropy* and *Interest focus*, which 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 *interest focus* (F_i) describes the proportion of the audience that are interested in i in a specific location. On the other hand, the *interest entropy* corresponds to the entropy measure of an interest distributed over the countries, which is calculated as:

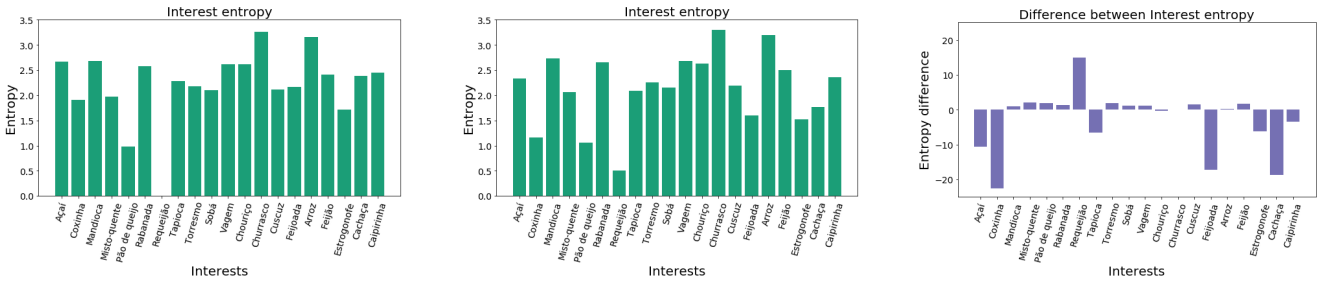
$$H_i = -\sum_{p=1}^{M} F_i \log_2(F_i), \text{ where } F_i = \frac{A_p(i)}{\sum_{p=1}^{M} A_p(i)}. \quad (3)$$

These metrics provide support for evaluating the spread of various interests around the globe. Interests that are very concentrated in a region tend to manifest a low *interest entropy* and a high *interest focus* in 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.

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 Brazil is taken into account. When we consider Brazil, the entropy decreases if the interest is significantly 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* decreases when Brazil is considered. If we use this approach to select typical Brazilian dishes, then 7 dishes meet this criterion: "Açaí", "Coxinha", "Tapioca", "Feijoada", "Estrogonofe", "Cachaça" and "Caipirinha". All of them are also selected by the 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", seem 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. For "Pão de Queijo", Facebook users from both 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 by both approaches, only the results considering the Brazilian typical dishes,



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

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

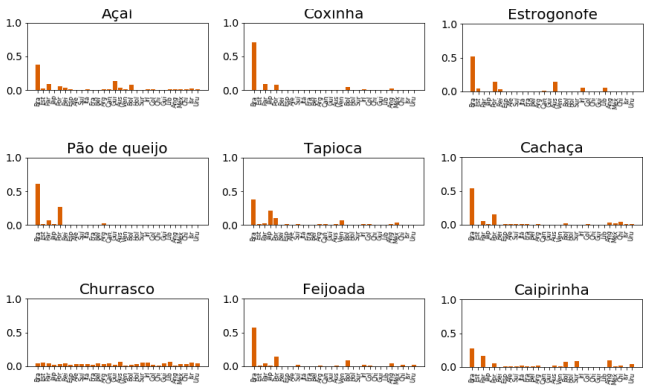


Figure 6: Interest focus considering Brazil.

according to the *interest entropy* difference, are reported as follows. Table 3 shows correlations between the *Immigrant 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 using the `Weighted tau` correlation, the Cultural distance ranking calculated by the cosine distance 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 dist.	0.38	0.05 (0.72)	0.06 (0.77)	0.33
Cosine dist.	**0.39**	-0.1 (0.47)	-0.14 (0.48)	0.33
Mean Abs. Error	0.39	-0.08 (0.54)	-0.15 (0.45)	**0.43**
Relative Error	0.39	**0.05** (0.7)	**0.06** (0.76)	0.33
Eart Mover's dist.	0.39	-0.08 (0.54)	-0.15 (0.45)	**0.43**

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. The goal of this study is to explore specific cultural attributes in order to develop a measure of distance that would most accurately

Rankings	WT$^{(i)}$	KT$^{(ii)}$	S$^{(iii)}$	J$^{(iv)}$
Euclidean dist.	**0.2**	**0.19** (0.16)	**0.27** (0.16)	0.33
Cosine dist.	0.17	-0.04 (0.78)	-0.03 (0.88)	0.33
Mean Abs. Error	0.18	0.0 (0.99)	-0.01 (0.97)	0.33
Relative Error	0.18	-0.01 (0.93)	0.00 (0.99)	**0.43**
Earth Mover's dist.	0.18	0.0 (0.99)	-0.01 (0.97)	0.33

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

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 previous research that employs various measures of distance in order to explain international migration patterns. The methodology we developed helps one understand the study of cultural attraction from the social media perspective. In addition, the cultural distance between countries can also be included as one additional attribute in classic gravity-type models of migration [5].

Cultural distance can influence migration flows since cultural aspects can be a factor of attraction between populations. On the other hand, immigration can change the culture of a country. The cultural measure we developed can be included as one additional variable in gravity-type models to explain flows of people around the world. Our paper focuses on how Facebook data can be correlated with migration processes and how a foreign culture, in our case, the Brazilian culture, is spread around the world.

Facebook data revealed to be reliable proxy to study international migration. In this paper, 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 greater than 0.7. Also, this data can 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. Our results suggest the cultural similarity between Brazil and some countries occur due to aspects such as geographic proximity (e.g. Paraguay, Argentina), linguistic similarity (e.g. Portugal, Angola), and most important, 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] Matheus Araujo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. 2017. Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *Proceedings of the 2017 ACM on Web Science Conference*. 253–257.

[3] Jean Boutaud, Anda Becuţ, 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

[4] 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

[5] 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

[6] 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

[7] Antoine Dubois, Emilio Zagheni, Kiran Garimella, and Ingmar Weber. 2018. Studying Migrant Assimilation Through Facebook Interests. In *Social Informatics*, Steffen Staab, Olessia Koltsova, and Dmitry I. Ignatov (Eds.). Springer International Publishing, Cham, 51–60.

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

[9] 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

[10] 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

[11] 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

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

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

[14] 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

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

[16] 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

[17] 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.

[18] 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

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

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

[21] Filipe N. Ribeiro, Lucas Henrique, Fabrício Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P. Gummadi. 2018. Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. In *Twelfth AAAI Conference on Web and Social Media (ICWSM'18)*. Stanford, USA.

[22] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabrício Benevenuto Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. 2019. On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook. In *ACM Conference on Fairness, Accountability, and Transparency (FAT*'19)*. Atlanta, USA.

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

[24] 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*.

[25] 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).

[26] 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

[27] 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.