Coimbra Vieira, Carolina; Lohmann, Sophie; Zagheni, Emilio

The Value of Cultural Similarity for Predicting Migration: Evidence from Food and Drink Interests in Digital Trace Data

Population and Development Review

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Coimbra Vieira, Carolina; Lohmann, Sophie; Zagheni, Emilio (2024) : The Value of Cultural Similarity for Predicting Migration: Evidence from Food and Drink Interests in Digital Trace Data, Population and Development Review, ISSN 1728-4457, Wiley, Hoboken, NJ, Vol. 50, Iss. 1, pp. 149-176,
https://doi.org/10.1111/padr.12607

The Value of Cultural Similarity for Predicting Migration: Evidence from Food and Drink Interests in Digital Trace Data

Carolina Coimbra Vieira, Sophie Lohmann and Emilio Zagheni

One of the strongest empirical regularities in spatial demography is that flows of migrants are positively associated with population stocks at origin and destination and are inversely related to distance. This pattern was formalized into what are known as gravity models of migration. Traditionally, distance is measured geographically, but other measures of distance, such as cultural distance, are also relevant in explaining migration flows. However, measures of cultural distance are not widely adopted in the literature on modeling migration flows, partially because of the difficulties associated with operationalizing and producing these measures across space and time. In this paper, we use a scalable approach to obtain proxies for measuring cultural similarity between countries by using Facebook data and illustrate the impact of incorporating these measures, based on food and drink interests, into gravity models for predicting migration. Our results show that, despite their limitations, the new measures of cultural similarity derived from Facebook data improve the prediction power of traditional gravity models and have a predictive capacity comparable to that of classic variables used in the literature, such as shared language and history. The results open up new opportunities for understanding the determinants of migration and for predicting migration when considering broader and complementary perspectives on the meaning and measurement of distance.

Introduction

One of the strongest empirical regularities in spatial demography is that flows of migrants are positively associated with population size at origin and destination and are inversely related to distance. This pattern was observed in the 19th century by Ravenstein (1889) and was later formalized by

Carolina Coimbra Vieira, Sophie Lohmann and Emilio Zagheni, Max Planck Institute for Demographic Research, 18057, Rostock, Germany. E-mail: carolcoimbra.dcc@gmail.com.

Zipf (1946) into what are known as gravity models of migration. Traditionally, distance is measured geographically. However, other measures, including those based on economic and cultural factors, have also been found to be relevant for explaining migration flows (Anderson 2011; Esses 2018; Caragliu et al. 2013; Böhme, Gröger, and Stöhr 2020; Lewer and Van den Berg 2008).

The cultural distance between two countries could, therefore, be a valuable predictor of migration flows given the bidirectional relationship between culture and migration. For instance, the cultural fit in terms of language, norms, and values is an important factor that people consider before moving between countries (Caragliu et al. 2013; Pedersen, Pytlikova, and Smith 2004). After moving, migrants then transmit cultural elements, such as food habits (Opare-Obisaw et al. 2000), from their origin country to their destination country and back again (Mesoudi 2018).

Measures of cultural distance are difficult to estimate and thus have not yet been widely adopted in gravity models for assessing and predicting migration. The few studies that have examined the impact of cultural dimensions or cultural distance on migration flows have typically relied on survey responses regarding norms, values, and beliefs, such as those from the World Values Survey (WVS; Inglehart 1997). Immigration data from Denmark, Germany, and the Netherlands illustrate that greater cultural distance, as derived from the WVS, is associated with less long-term mobility (White 2013). Overall, cultural distance, as derived from the WVS, seems to play an important role in predicting migration flows between European countries (Caragliu et al. 2013). However, survey approaches that focus on migration suffer from significant limitations, such as the difficulty of reaching migrants, and the complexity and high costs associated with running a cross-national survey with a migration focus.

In this paper, we use complementary measures of cultural similarity based on cultural norms, values, and beliefs derived from surveys (i.e., the WVS) for the study of migration flows. We expand the analysis of the impact of culture on migration flows by adding measures of cultural similarity based on cultural attributes regarding food and drink interests derived from social media data (i.e., Foursquare and Facebook). This article cannot distinguish between all the mechanisms underlying the complex relationship between culture and migration, and the aim of this study is not to establish a causal link between them. We do, however, demonstrate that culture is an important aspect to consider when studying migration and that the inclusion of measures of cultural similarities improves predictions of migration flows, even after accounting for classic predictors of migration. We expand the literature by showing the impact of adding measures of cultural similarity derived from social media data based on food and drink interests (i.e., food and drink similarity) to the analysis of migration flows.

Food and drink are two of the most basic needs of human beings. The manner in which people interact with food, from the procurement and selection of food to its preparation and consumption, reflects complex interrelationships and interactions among individuals, the society in which they live, and their culture (Axelson 1986; Ferguson, Iturbide, and Raffaelli 2020). Food studies have become an important interdisciplinary field of study that focuses on the relationship between food and human experience and the relationships between food, culture, and society (Almerico 2014). Food production, distribution, and consumption are all shaped by cultural codes (Counihan and Van Esterik 2012) and represent cultural acts (Montanari 2006). Food preparation[1] is one of the topics included in the Lists of Intangible Cultural Heritage provided by UNESCO,[2] which covers cultural practices and expressions of intangible heritage. More generally, food can be seen as an important marker of cultural identity (Kittler, Sucher, and Nelms 2016). Out of all categories of interests on Facebook (e.g., food and drink, news and entertainment, hobbies and activities, sports and outdoors), food and drink are the only interests that belong to a universal category, given that food and drink are two of the most basic needs of human beings. Some interests are specific to certain demographic groups; for instance, not everyone is interested in sports or celebrities. By contrast, food is popular across a wide demographic spectrum. In this context, given the importance of food to culture (Ashley et al. 2004; De Solier and Duruz 2013; Recchi and Favell 2019), we consider measures of cultural similarity based on food and drink interests that are derived from social media data.

We propose the use of measures of food and drink similarity developed by Vieira et al. (2022) and evaluate their potential in predicting migration. Unlike survey data that need to rely on different rounds of survey to be collected, these measures are timely, cost-effective, and scalable as they are based on aggregate data from Facebook that are freely and publicly available through the Facebook Advertising Platform (we will refer to these data as Facebook Ads data). We illustrate the applicability of the proposed approach by showing how these new measures of food and drink similarity can be used to predict migration and explain migration flows. The measures of food and drink similarity derived from Facebook Ads have, despite their limitations, a capacity to predict migration flows that is comparable to that of classic variables used in the literature to represent the cultural dimension, such as shared language and shared history. Additionally, the measures of food and drink similarity derived from Facebook Ads are able to capture changes quickly, especially when migration patterns change rapidly due to crises. In this context, we expect that these measures of food and drink similarity derived from Facebook Ads could represent, almost in real-time, the cultural changes that occur during big and unexpected migration events (e.g., the migration of Ukrainians after Russia's invasion). Furthermore, the Facebook Ads measures of food and drink similarity introduce a more nuanced view of symmetric and nonsymmetric measures of similarity, opening

up new opportunities for predicting and understanding the determinants of migrations.

Background

Cultural distance measures operational parameters that can be used as proxies for cultural dimensions. They allow researchers to estimate the extent to which countries differ culturally (Tung and Verbeke 2010). The cultural dimensions used to measure culture can vary depending on the focus of the research (Mohr et al., 2019). For instance, the study of culture can focus on aspects of daily life by considering cultural objects, such as the clothes people wear, the music they listen to, and the food they eat (Recchi and Favell 2019; Kwantes and Glazer 2017). Food studies is an important interdisciplinary field that recognizes food as a central aspect of acculturation, cultural practices, and cultural identity (Ferguson, Iturbide, and Raffaelli 2020; Montanari 2006; Ashley et al. 2004; De Solier and Duruz 2013; Kittler, Sucher, and Nelms 2016).

Operationally, culture has been traditionally measured in terms of norms, values, and beliefs via sampling surveys (Kwantes and Glazer 2017) in which the survey responses are used to characterize cultural aspects of a country (e.g., Schwartz's value survey (Schwartz 1994), the WVS (Inglehart 1997), and Hofstede's[3] cultural characteristics (Hofstede 1983)) and to evaluate the relative distance between countries (Gupta, Hanges, and Dorfman 2002; De Santis, Maltagliati, and Salvini 2016; Mucciardi and De Santis 2017; Muthukrishna et al. 2020).

Such studies based on surveys are highly valuable but also have important limitations. In addition to measurement error (Groves and Lyberg 2010), the results may suffer from various biases (Suchman 1962), like social desirability bias, question order bias, and acquiescence bias. Furthermore, surveys are costly and require a long time to run. For example, most government statistics are updated only once a year (often with a delay), and major surveys such as the European Values Study (EVS) and the WVS are often spaced even further apart (the EVS is conducted once every nine years, and the WVS is carried out once every five years). This lack of timely information makes it difficult for decision-makers to respond dynamically to shifting circumstances. While all demographic studies involve uncertainty and complexity, migration predictions are particularly uncertain (Bijak and Bijak 2022). For instance, migration flows related to refugee movements[4] are among the most volatile forms of migration and are, therefore, the most difficult to predict. These dynamic shifts in migration flows illustrate the need for more frequent data to track migration flows and improve predictions. To overcome some of these limitations, we propose an approach that relies on passively collected data from social media, which can be used to complement data from existing sources.

Social media advertising platforms provide complementary tools that can be used to measure cultural preferences and that allow for comparisons across regions via passively collected data (You et al. 2017). As one of the first studies to address this question using online data sources, Silva et al. (2014) identified cultural boundaries and similarities across populations by clustering them based on the analysis of food and drink habits. However, their analyses of culinary habits around the world were limited to Foursquare check-ins, which considered only 101 categories and thus underestimated the variety of users' interests.

In one of the first studies on this topic using Facebook Ads data, Vieira et al. (2020) examined the similarities between selected countries and Brazil based on their population's interests in typical Brazilian dishes. However, the results were limited to dishes listed on Wikipedia, which restricted the potential list of dishes. Moreover, because some countries do not have a Wikipedia page dedicated to listing their typical dishes, the methodology was not scalable. Obradovich et al. (2022) also used data from Facebook Ads to examine cross-national cultural differences across nearly 60,000 interests. They validated their work by comparing the cultural distances calculated using their measurements with those of traditional survey-based measures. However, as the authors included a wide range of cultural features, from politics to national parks, and used a "black box" model to represent countries' cultures, it is hard to assess exactly what their index was measuring. More recently, Vieira et al. (2022) presented a scalable, data-driven methodology for measuring the cultural similarities between countries based on the most popular food and drink in each country from a list containing more than 200,000 interests on Facebook Ads (Speicher et al. 2018). Relative to Obradovich et al. (2022), the methodology proposed by Vieira et al. (2022) compared countries using fewer, but explicitly known attributes selected from a very large data set. In other words, interests that were not relevant to any of the countries were disregarded to reduce feature sparsity. They presented two measures of cultural similarity, including the first asymmetric measure of cultural similarity derived from social media data.

The literature that we just summarized suggested methodologies based on social media data to measure cultural similarity between countries and then correlated these measures with survey-based measures. To the best of our knowledge, ours is the first paper to evaluate how suitable the use of an asymmetric measure of similarity is for predicting migration. In this work, we decided to evaluate the impact of adding these measures of cultural similarity to a gravity model to predict migration. In order to test the Facebook measures against the most stringent baseline possible, we compared its predictive capacity with that of measures of cultural similarity derived from the WVS (Inglehart 1997) and Foursquare data (Silva et al. 2014).

Facebook Ads data have become an important tool in demographic research, especially for studying migration patterns (Leasure et al. 2023; Zagheni, Weber, and Gummadi 2017; Dubois et al. 2018; Spyratos et al. 2019; Alexander, Polimis, and Zagheni 2019; Palotti et al. 2020). However, the study of international migration and the development of models to explain and predict flows of people between countries are not new (Massey et al. 1993). One of the most traditional prediction approaches is based on gravity-type models (Tinbergen 1962; Lewer and Van den Berg 2008; Cohen et al. 2008; Ramos 2016). For example, Cohen et al. (2008) developed an algorithm to project future numbers of international migrants from any country or region to any other country or region. The model considers the population and the geographical area of the origin and the destination country and the geographic distance between the origin and destination. Subsequently, researchers have added to this model by identifying other variables, such as social variables (e.g., mortality rate) (Kim and Cohen 2010); historical variables, such as shared history and shared language (Lewer and Van den Berg 2008; Beine, Bertoli, and Moraga 2016; Kim and Cohen 2010; Caragliu et al. 2013; Abel, Raymer, and Guan 2019); and online search keywords (Böhme, Gröger, and Stöhr 2020). For instance, Böhme, Gröger, and Stöhr (2020) showed how geo-referenced online search data can be used to measure migration intentions in origin countries and to predict bilateral migration flows. Moreover, distance measures that go beyond estimating geographic distance to, for example, assess administrative, political, economic, or cultural distance (Ghemawat 2001) are important variables that should be considered by migration prediction models.

Most of the existing studies that analyzed cultural changes in relation to migration were restricted to one or a few countries, or, if they took a broader international perspective, used cultural distance measures that were symmetric by construction (Rapoport, Sardoschau, and Silve 2020). Since migration is neither homogeneous across countries nor symmetric, we apply an asymmetric measure of cultural similarity across many countries in order to more accurately represent processes of international cultural exchange.

Data

In this section, we describe the main data sources we used to collect international data for our prediction models. We present the data sources we used to measure the cultural and food and drink similarity between countries: the WVS (Inglehart 1997), Foursquare (Silva et al. 2014), and Facebook Ads (Vieira et al. 2022). We also provide a description of the data sources for gravity model variables such as population, area, and geographic distance, as well as for migration flows as the outcome variable. To ensure the comparability of our results with previously suggested indices of similarity

based on Foursquare data, our analysis focuses on a subset of 16 of the most popular countries by number of Foursquare check-ins (Silva et al. 2014).

The countries selected for the analysis were chosen based on a compromise across three different criteria. First, we wanted to match the list of countries selected by Silva et al. (2014) to allow for comparisons between the previous literature using Foursquare data and our results using Facebook data. Second, the countries we chose cover a large portion of geographic areas and populations across the world. Finally, and importantly, we selected countries with high Facebook penetration rates: Argentina, Australia, Brazil, Chile, Great Britain, France, Indonesia, Japan, South Korea, Malaysia, Mexico, Russia, Singapore, Spain, Turkey, and the United States. In other words, we favored a choice of countries with comparatively low and consistent biases over a broader selection of countries that would be more heterogeneous in terms of biases and for which the interpretation of results would be more complex.

Facebook ads data

Vieira et al. (2022) collected data regarding Facebook users' interests in food and drink and proposed measures of cultural similarity between countries. The data collected from Facebook Ads refer to the number of Facebook monthly active users (i.e., active over the past 30 days) who matched the demographic attributes targeted at the time of data collection. The Facebook Marketing API enables marketers and researchers to estimate the monthly active Facebook user count for a proposed advertisement, aligning with specified input criteria (Kosinski et al. 2015). The platform provides a set of customizable demographic attributes, such as age, gender, home location, and interests, allowing advertisers to tailor their input queries. Attributes like age, gender, and location are explicitly declared by the users in their profiles, whereas interests can be either declared by the user or inferred by Facebook based on user activities such as posting or interacting with content (e.g., liking content, sharing content, or updating one's status). The methodology proposed by Vieira et al. (2022) consisted of selecting a subset of popular foods and drinks for each country and then creating a vector representation according to Facebook users' interests in those foods and drinks. Finally, they measured the similarity between those country-level vectors. Methodological details are available in Vieira et al. (2022).[5] Measures derived from the Facebook Ads data, including the code used to analyze the data sets and generate the figures, are available in a public web repository.[6]

Two measures of similarity were proposed by Vieira et al. (2022)—Facebook asymmetric similarity and Facebook symmetric similarity—depending on the subset of food and drink used to create the vector representations. The *asymmetric similarity* between two countries, c_1 and c_2, is measured in terms of the most popular food and drink in c_1, whereas the

similarity between c_2 and c_1 is measured in terms of the most popular food and drink in c_2. In this case, since the similarity between c_1 and c_2 is different from the similarity between c_2 and c_1, this measure is not symmetric.

However, we could also measure the similarity between two countries by considering a fixed set of interests for both countries. In this case, we can refer to the measure as *symmetric similarity*, corresponding to the measure of similarity between two countries considering the union of the most popular food and drink in these countries. Since the subset of interests is fixed, the similarity between c_1 and c_2 is equal to the similarity between c_2 and c_1. In our models, we refer to the Facebook asymmetric measure of similarity as Facebook asymmetric similarity—food origin or food destination—depending on which subset of top food and drink, from the country of origin or the country of destination, we considered in the measure of similarity. We refer to the symmetric measure of similarity as Facebook symmetric similarity.

For the Facebook measures of food and drink similarity (Vieira et al. 2022), selected the top 50[7] types of food and drink in each country. In this case, the asymmetric measure of similarity between two countries, c_1 and c_2, corresponds to the cosine similarity between the 50-dimensional vector representation of each country in terms of the 50 top foods and drinks in country c_1. The symmetric measure of similarity between two countries, on the other hand, is given by the cosine similarity between the vector representation of each country in terms of the 394 foods and drinks. The set of 394 interests corresponds to the union of the top 50 interests in each of the 16 countries.

The main aim of this paper is to assess the extent to which the considered Facebook measures of cultural similarity—using only food and drink as cultural markers—are meaningful predictors of migration flows. In this sense, it is important to validate our results obtained with the Facebook Ads data and to ensure their comparability with the results of prior research. We selected the two most relevant data sets for comparing measures of cultural similarity: the WVS and Foursquare. The WVS is an established and traditional data set based on large-scale representative survey data along several cultural dimensions reflecting cultural norms, values, and beliefs. The Foursquare data set (Silva et al. 2014) is based on data on users' food and drink habits collected from Foursquare check-ins. Although the measures of cultural similarity from the WVS and the Foursquare data focus on different aspects of culture, both measures are used as baselines for the Facebook measures of food and drink similarity. However, as mentioned before, both data sets have significant limitations. The main disadvantages of the WVS are the costs and the operational time needed to release new survey waves, which have typically been conducted for about five years. The Foursquare platform, in contrast, is not as widely used as Facebook. In addition, the platform is heavily biased from a demographic point of view

(e.g., young men are more likely to use Foursquare than women or older people,[8,9] and most users are from the United States,[10] although only 2% of the US population use Foursquare[11]) and is limited to people who explicitly share their locations when visiting a place (i.e., check-in). Moreover, access to the Foursquare API is not free.[12]

World Values Survey data

The WVS data set considers several cultural dimensions, such as religion, politics, economics, and lifestyle. Inglehart (1997) identified two major dimensions derived from the WVS and proposed a cultural map[13] of the world in which the location of each country is given by the scores on these two dimensions. Based on the location of each country in the WVS cultural map from 2020, each country was represented by a vector of two dimensions corresponding to the two dimensions of the WVS cultural map: traditional values (which emphasize the importance of religion, parent–child ties, deference to authority, and traditional family values) versus secular-rational values (represented by societies that place less emphasis on religion, traditional family values, and authority); and survival values (emphasis on economic and physical security) versus self-expression values (emphasis on environmental protection, equality, and rising demands for participation in decision-making in economic and political life). Then, we measured the cosine similarity between each pair of countries to obtain the WVS similarity.

The measure of cultural similarity derived from the WVS data is a representation of a more traditional measure of culture based on norms, values, and beliefs. The selection of the WVS as a data source for measuring cultural similarities across countries is based on two main criteria: coverage and wave updates. The WVS, which is one of the most authoritative and widely used cross-national surveys in the social sciences, has an extensive geographical and thematic scope and is conducted globally[14] every five years. In addition, there is a strong association between the cultural dimensions from the WVS cultural map and other cultural dimensions derived from surveys (Kaasa and Minkov 2022; Taras, Rowney, and Steel 2009).

Foursquare data

Silva et al. (2014) identified cultural boundaries and similarities across populations by clustering them based on the analysis of food and drink habits via Foursquare check-ins. They also proposed a cultural map in which the location of each country is given by the two first principal components after applying the principal component analysis algorithm over a high dimensional preference vector based on Foursquare check-ins in different subcategories of bars and restaurants. Based on the location of each country in

the cultural map proposed by Silva et al. (2014), we measured the cosine similarity between them to obtain the Foursquare similarity.

We compared different measures of similarity derived from Euclidean and cosine distance. The two measures are highly correlated with each other (WVS data (0.64) and Foursquare data (0.89); see Figures A2 and A3 in the online Appendix). We did not observe substantive changes in our model when using cultural similarities derived from Euclidean or cosine similarity. For consistency purposes, all the measures of cultural similarity derived from the WVS, Foursquare, and Facebook Ads data are based on cosine similarity.

United Nations data

The population size of each country in 2019[15] comes from the United Nations. The data include estimates of the total population for all countries and are made available in the 2019 revision of the World Population Prospects. We also collected the estimates of the number of international migrants and the migrant stocks from each country of origin in each country of destination for 2019 from the United Nations website.[16]

CEPII GeoDist data

GeoDist (Mayer and Zignago 2011) makes available the exhaustive set of gravity variables developed by Mayer et al. (2005) to analyze market access difficulties in global and regional trade flows. The data set incorporates country-specific geographical variables for the world's countries, including the area of each country in square kilometers (km^2). Moreover, the data set includes variables that apply to pairs of countries, from which we used the variables of geographic distance and shared history. The geographic distance between each pair of countries is based on bilateral distances between the biggest cities of those two countries, weighted by the share of the city in the country's overall population (Mayer and Zignago 2011). Finally, shared history indicates whether the two countries have ever had a colonial link. Based on the information from the data set, a colonial link is established if the two countries have had a common colonizer after 1945, have ever had a colonial link, have had a colonial relationship after 1945, are currently in a colonial relationship, or were/are the same country.

CEPII Language data

The CEPII Language data set (Melitz and Toubal 2014) provides separate measures of common native language, common spoken language, common official language, and linguistic proximity between different native languages. In our model, we use an indicator of whether the two countries share a common official language as well as an indicator of the linguistic

proximity between the two countries' languages. The first indicator is a binary variable that codes whether the two countries share at least one official language. The second indicator calculates the linguistic proximity on the basis of the Ethnologue classification of language trees between trees, branches, and subbranches (Fearon 2003; Laitin 2000). This variable can take four values: zero for languages belonging to separate family trees; 0.25 for languages belonging to different branches of the same family tree (e.g., English and French); 0.50 for languages belonging to the same branch (e.g., English and German); and 0.75 for languages belonging to the same subbranch (e.g., German and Dutch).

World Bank DATA

The gross domestic product per capita, or GDP per capita (constant 2010 US$), of each country in 2019 was collected from the World Bank.[17]

Migration flow data

Finally, the dependent variable of our models is the migration flow by the origin country and the destination country. Due to the lack of migration flow data at the global level, we use the estimated values from the demographic accounting, a pseudo-Bayesian approach proposed by Abel and Cohen (2019). The estimations are available only for five-year bilateral migration flows. For our analysis, we use the latest estimates of migration flows for the 2015–2019 period.

Gravity models to predict migration

In order to investigate whether the measures of cultural similarity can improve the prediction of migration flows, we tested different models inspired by traditional gravity models (Cohen et al. 2008). Although they have their limitations (Beyer, Schewe, and Lotze-Campen 2022), gravity models are among the most traditional models used to predict migration flows. The gravity model is a log-linear model,[18] as shown by Equation (1). The classic model uses only the population of origin and destination (P_o, P_d), the area of the country of origin and destination (A_o, A_d), and the geographic distance between origin and destination ($D_{o,d}$) as independent variables to predict migration flows ($M_{o,d}$) between the country of origin and the country of destination:

$$\log_{10}\left(M_{o,d}\right) = \beta_0 + \beta_1 \log_{10}\left(P_o\right) + \beta_2 \log_{10}\left(A_o\right) + \beta_3 \log_{10}\left(P_d\right) + \beta_4 \log_{10}\left(A_d\right) + \beta_5 \log_{10}\left(D_{o,d}\right) + \epsilon_{o,d}. \quad (1)$$

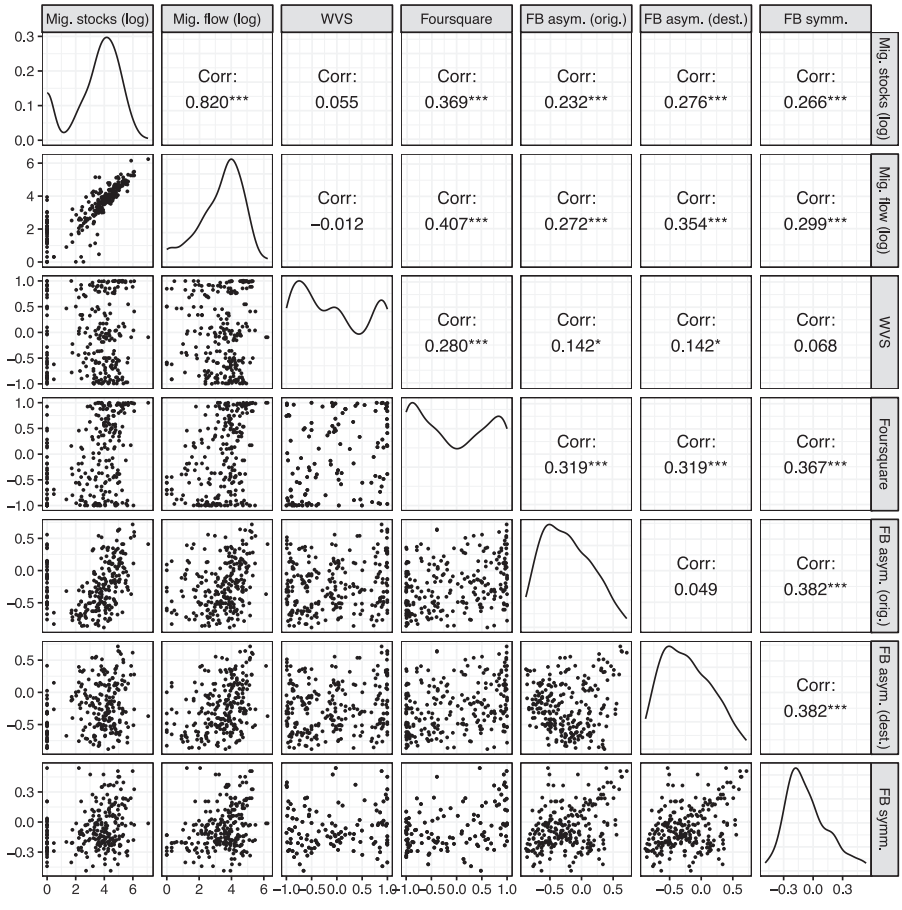
We added additional independent variables that might promote or deter migration based on prior literature that has identified them as relevant

predictors (Anderson 2011). This allowed us to test the extent to which measures of culture affect the predictive capacity of the model beyond the more stringent baseline. We examined a series of gravity models. The first model (Model 1) refers to the classic version of a gravity model (Cohen et al. 2008), represented by Equation 1, plus the GDP per capita of both the origin and the destination country, and the migrant stocks between the origin and the destination country. All the independent variables included in Model 1 are referred to as basic variables, as they are important predictors that are commonly added to gravity models (Tinbergen 1962; Cohen et al. 2008; Böhme, Gröger, and Stöhr 2020). Given the importance of having a common language and a common colonial history (Beine, Bertoli, and Moraga 2016; Kim and Cohen 2010; Lewer and Van den Berg 2008; Caragliu et al. 2013; Abel, Raymer, and Guan 2019), the second model (Model 2) builds on the first model and adds variables related to shared language and shared history. Shared language and history (e.g., colonial history) are often included in gravity models as cultural variables. However, other types of cultural differences may also be relevant for predicting and explaining migration. For instance, differences in countries' cultural norms, values, and beliefs can affect migration flows (Caragliu et al. 2013). To take into account the impact on migration of the cultural norms, values, and beliefs in a country (Esses 2018; Caragliu et al. 2013), we specified the third model (Model 3) by adding the WVS cultural similarity measure to the second model.

Finally, we assessed the impact of adding measures of food and drink similarity derived from social media data to each of the three models presented. For instance, Model 2 + FB asymmetric adds the first asymmetric measure of similarity using data from Facebook Ads to Model 2. The Facebook asymmetric measure of similarity consists of two variables, Facebook asymmetric similarity (food origin) and Facebook asymmetric similarity (food destination), which represent measures of similarity based on the food and drink that are popular in the country of origin and the food and drink that are popular in the country of destination. Model 2 + FB symmetric and Model 2 + Foursquare add, respectively, the symmetric measure of similarity derived from Facebook Ads data and the measure of similarity derived from Foursquare data to Model 2. Each measure of cultural similarity derived from social media data was added separately to the models to assess their individual impact in predicting migration flows. Due to the relatively high correlation between the Facebook measures of similarity (0.38) and between the Foursquare and Facebook measures of similarity (0.32 and 0.37), as shown in Figure 1, we only tested those measures separately to reduce potential issues of collinearity.

In all these models, the dependent variable is the logarithm of the estimated migrant flow in 2019. In order to calculate the logarithm of the dependent variable, we added an offset (equal to one) to all the observations. The resulting coefficients and statistics for each of these models are

FIGURE 1 Distribution and correlations between the measures of similarity and migration flow (in the logarithm scale) between countries. Each dot represents a pair of countries within the 16 countries we analyzed



***p <0.001; **p < 0.01; *p < 0.05.

presented in Table T1 in the online Appendix. In this table, the columns represent the models and the rows describe each of the variables in the prediction model.

Cross-validation

To properly evaluate the models and predictions, while avoiding overfitting, we decided to evaluate each model against a test sample of the data that was not seen by the model during the fitting phase that relied on the training data. Model fitting and evaluation were done only on the training data. The evaluation of predictive accuracy or errors was done on the testing data.

TABLE 1 Overall prediction errors for each model evaluated using cross-validation

Models	RMSE	*R*-squared	MAE	WAIC*
Model 1: area, population, distance, GDP per capita, migrant stocks	0.61	0.78	0.46	440.05
Model 1 + FB asymmetric	0.59	0.79	0.45	430.40
Model 1 + FB symmetric	0.61	0.78	0.46	441.28
Model 1 + Foursquare similarity	0.61	0.78	0.46	441.86
Model 2: Model 1 + shared language and history	0.58	0.80	0.45	419.91
Model 2 + FB asymmetric	0.59	0.80	0.45	422.42
Model 2 + FB symmetric	0.58	0.80	0.45	420.06
Model 2 + Foursquare	0.58	0.80	0.45	421.31
Model 3: Model 2 + WVS	0.55	0.82	0.43	389.93
Model 3 + FB asymmetric	0.55	0.82	0.43	391.61
Model 3 + FB symmetric	0.55	0.82	0.43	392.43
Model 3 + Foursquare	0.55	0.82	0.43	389.01

Abbreviations: MAE, mean absolute error; RMSE, root mean squared error; WAIC, Watanabe–Akaike or widely applicable information criterion.
*Metric applied just to the final model using the full input data set (240 pairs of countries).

For this evaluation, we used the leave-one-out cross-validation (LOOCV) approach.

The LOOCV approach requires one model to be evaluated for each point in the training data set. Given a data set with N data points (in our case, pairs of countries), the cross-validation works in the following way: (i) the model is trained on $N - 1$ data points; (ii) the model is tested against the one data point that was left out in the previous step; (iii) the prediction error is calculated; (iv) the first three steps are repeated until the model is trained and tested on all data points; and (v) the overall prediction error is generated (e.g., by taking the average of the prediction errors across all models).

Table 1 shows the average prediction errors for three different measures for each model. The rows represent the models and the first three columns represent the three measures of prediction errors. The root mean squared error (RMSE) is a proxy for the average difference between the predictions made by the model and the actual observations. The lower the RMSE, the more closely a model can predict the actual observations. The mean absolute error (MAE) is the average absolute difference between the predictions made by the model and the actual observations. The lower the MAE, the more closely a model predicts the actual observations. Finally, the *R*-squared is a measure of the correlation between the predictions made by the model and the actual observations. The higher the *R*-squared, the more variance in the data is explained by the model. Each of the three metrics provided in the output gives us an idea of how well the model performed on previously unseen data. However, the *R*-squared measure itself

is not reliable, since more variables always increase the metric, even if new variables are only marginally predictive. To address this issue, we included other measures that penalize the number of variables in their calculation. The last column in Table 1 shows the Watanabe–Akaike or widely applicable information criterion (WAIC) (Gelman et al. 1995) for each of the models tested. We used the full input data set (without cross-validation), which consists of 240 pairs of countries. The lower the Watanabe–Akaike, the more closely a model can predict the actual observations. Table T1 also shows the adjusted *R*-squared and the resulting coefficients and statistics for each of these models. In this table, the columns represent the models and the rows display each of the variables in the prediction model. In the next section, the resulting coefficients and statistics from Table 1 and Table T1 are described in more detail.

Results

Table T1 shows in detail all the coefficients for each of the variables included in the gravity models that we tested using the full input data set corresponding to 240 pairs of countries. Table 1 shows the results averaged across cross-validations, except for the WAIC, which was calculated from the model using the full input data set. To evaluate the impact of adding measures of cultural similarity to the migration model, we first assess the correlation between the variables considered.

Figure 1 shows the correlation between each of the measures of similarity and migration flows (in the logarithm scale) between each pair of countries within the 16 countries we analyzed. We observed that the symmetric and asymmetric measures of food and drink similarity derived from Facebook Ads data showed a positive correlation (0.38). Similarly, the measures of food and drink similarity based on Foursquare data and Facebook Ads data were also positively correlated (0.37 between the Foursquare measure and the Facebook symmetric measure and 0.32 between the Foursquare measure and the Facebook asymmetric measure). The Foursquare and Facebook measures of food and drink similarity were highly correlated with each other and captured similar patterns of food and drink interests across countries. Next, we compared the cultural similarities derived from social media with those derived from the WVS. Although cultural similarities based on the WVS data and the Facebook symmetric measure did not capture the same cultural attributes and were not substantially associated (0.07), the WVS data were positively correlated with both Facebook asymmetric (0.14) and Foursquare measures of food and drink similarity (0.28).

The measures of food and drink similarity derived from social media data did not exhibit a strong correlation with the metrics obtained from the survey data. Whereas the metrics derived from the WVS data encompassed

culture in terms of norms, values, and beliefs, the metrics derived from the Foursquare and Facebook data primarily emphasized the interest in food and drink as cultural markers. The differences in the nature of the data suggest that the WVS cultural similarity measure captured different aspects of culture that were not reflected by the interests in food and drink drawn from the social media data. The measures derived from the Foursquare and Facebook data, on the other hand, were highly correlated to each other and captured a similar pattern of food and drink interests across countries.

We observed that cultural similarity based on the WVS data was not positively correlated with migration flows (−0.01). This result means that countries that were close to each other in the WVS cultural map had slightly smaller migration flows between them. Table T1 shows a significant negative effect of the WVS cultural similarity on migration flows, meaning that a high WVS cultural similarity was associated with smaller migration flows. Despite the unexpected negative effect of the WVS cultural similarity on migration flows, the adjusted *R*-squared improved (0.83) when the WVS cultural similarity was added to the gravity model. This result confirms the importance of taking a country's cultural norms, values, and beliefs into account when fitting migration models (Esses 2018; Caragliu et al. 2013). In contrast, the measures derived from social media data all showed a positive correlation with migration flows (Foursquare 0.41; Facebook Ads asymmetric 0.27 and 0.35, Facebook Ads symmetric 0.3), which means that a high food and drink similarity was associated with larger migration flows.

Even with this stringent baseline of adding geographic and economic variables, we found that including measures of cultural similarity derived from Facebook data focusing on food and drink improved predictions beyond what could be achieved with all these other predictors. The coefficients from Model 1, including the Facebook measures of food and drink similarity, were statistically significant, and the predictive capacity of the model increased. This suggests that the Facebook measures of food and drink similarity are important predictors of migration, capture different patterns, and can be used to identify directional processes. This is not the case for the model that includes all the basic variables and shared language and history (Model 2). In other words, we did not observe improvements when adding the Facebook measures of food and drink similarity, which means that there is likely an overlap in the explanatory power of the measures of food and drink similarity derived from Facebook data and shared language and history. Figure A1 in the online Appendix shows the significant positive correlation between the measures of food and drink similarity derived from Facebook data and shared language and history. The estimated coefficients in Table T1 for the measures of cultural similarity based on interests in food and drink derived from Facebook data in Model 2 were not significant. Finally, even though the measures of cultural similarity derived from the WVS data and the social media data captured different aspects of culture, we did

not observe significant coefficients when we added them to the model that included all the basic variables and the WVS cultural similarity (Model 3). However, we observed a slight improvement in the adjusted *R*-squared compared to Model 3 when the measures derived from the Foursquare data and the asymmetric measure of food and drink similarity derived from the Facebook data were added to the model.

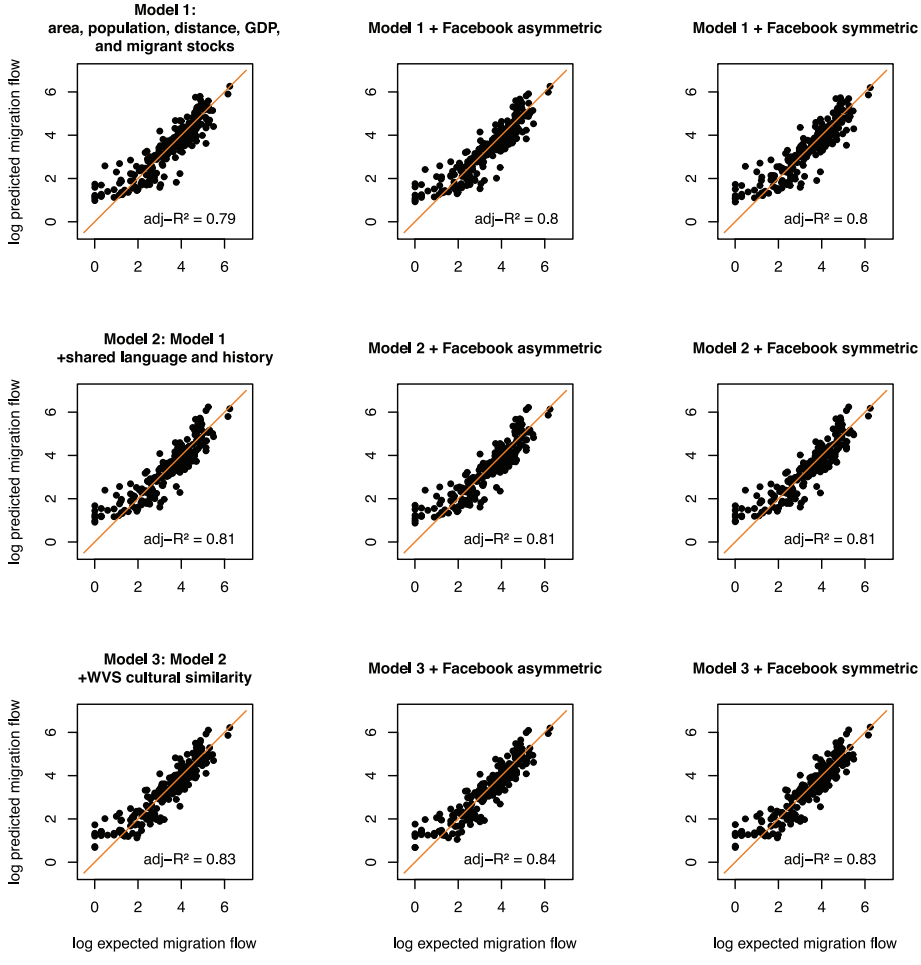
Despite the small improvement in the prediction of migration flows for the time point considered, the measures of food and drink similarity derived from the Facebook data had a predictive capacity comparable to that of the classic variables used in the literature, such as shared language and shared history. In addition, the measures of food and drink similarity derived from the Facebook data contributed to predictive models of migration by adding not just a timely but also an asymmetric component. The measures of similarity from the Facebook data were measuring country-level indices of cultural interests, which can shift precisely through migration. While systems of belief within a single culture (e.g., the majority culture in a country) should not change quickly, the ratio of the majority culture to the minority culture(s) can shift, leading to changes in country-level interests that would be observable in digital trace data. Particularly given the limitations of other measures of cultural similarities, the use of Facebook Ads data can provide an effective means of capturing such changes in a way that complements other measures.

Figure 2 shows a comparison between the expected migration flows estimated by Abel and Cohen (2019) and the migration flows predicted by each model. The orange line represents the expected distribution, where the predicted migration flow is equal to the expected migration flow. The distribution of the dots, which corresponds to pairs of countries, changes from one model to the other, and the predictions become closer to the expected values for migration flows. Overall, we observed that the baseline and the more traditional models overestimated migration flows for pairs of countries between which there was little migration, and underestimated migration flows for pairs of countries between which there were larger migration flows. This pattern became slightly less evident with the inclusion of other variables, including the measures of food and drink similarity derived from Facebook.

Discussion

We showed the impact of adding measures of cultural similarity, derived from both survey data and social media data, to gravity models in order to predict migration flows. Our results indicated that the measure of cultural similarity derived from the WVS data helped to improve migration prediction and that the measure of food and drink similarity derived from Foursquare data was highly correlated with migration flows. However,

FIGURE 2 Comparison between the expected migration flows (*x*-axis) and the migration flows predicted (*y*-axis) by each one of the models using the full input data set (240 pairs of countries). Both axes are on a logarithmic scale. Each dot represents a pair of countries within the 16 countries we analyzed



in terms of scalability and reproducibility, the use of these measures may have some disadvantages. As was mentioned before, surveys are costly and require substantial operational time. For example, the WVS is carried out every five years. The Foursquare data have a different set of limitations. In particular, the Foursquare platform is not as widely used as Facebook, and it is heavily biased from a demographic point of view. Moreover, the Foursquare data set we considered was over five years older than the Facebook Ads data and over six years older than the WVS data. During this period of time, significant cultural changes may have happened, given that the world is continuously changing in terms of connectivity across regions.

With more than 2.7 billion users worldwide,[19] Facebook captures a larger and more diverse population than other social media. Considering the availability of Facebook's data, our methodology could be easily scaled to consider more countries. Moreover, the data from Facebook Ads are freely available and can be continuously updated and collected, which makes this approach timely, cost-effective, reproducible, and scalable. Thus, the relevance of these types of analyses in traditionally data-poor contexts, like in low- and middle-income countries, will likely increase in the future.

Given the advantages of using Facebook Ads data, we provided a stringent test of the incremental effects of the Facebook measures, and our results showed that cultural similarity, as measured by food and drink interests, explained migration flows to an extent that was comparable to that of standard predictors such as shared language and shared history. Besides the advantages of using Facebook data to measure food and drink similarity, the approach we presented had additional advantages due to its use of an asymmetric measure of similarity. Most of the gravity models relied on symmetric variables to predict migration, which is itself an asymmetric phenomenon. Since the migration flows between countries are asymmetric (e.g., there are more Chileans in Spain than Spaniards in Chile), we would expect that the similarity in terms of food and drink interests would be asymmetric as well (e.g., there are more Chileans interested in Spanish food than Spaniards interested in Chilean food). This phenomenon was reflected in the coefficients of the model using the Facebook asymmetric similarity measure, which showed a stronger effect on the popularity of the destination country's dishes in the country of origin than vice versa. For example, Chileans' interest in Spanish dishes would be a stronger predictor of how many Chileans moved to Spain than Spaniards' interest in Chilean dishes. We found evidence of asymmetric patterns, such that the cultural markers in the country of destination were more closely associated with migration flows than the cultural markers in the country of origin. We, therefore, recommend that future research in this area take asymmetry into account when predicting migration.

To the best of our knowledge, our study is the first to propose a scalable, rapidly available, and asymmetric measure of similarity derived from social media data to predict migration. Our findings contribute to the literature by (i) showing the importance of cultural similarity, as derived from food and drink interests in social media data, for predicting migration; and (ii) allowing for rapid predictions of current migration flows ahead of the release of official statistics. For instance, Leasure et al. (2023) leveraged data from Facebook Ads to monitor in real-time subnational population sizes and internal displacement in Ukraine on a daily basis, disaggregated by age and sex. Similarly to Leasure et al. (2023)'s work, our methodology could capture rapid changes in populations' interests across countries, for instance, due to unexpected migration, and could help in predicting migration flows.

The primary objective of this study was to assess the value of examining cultural similarity when studying migration. Specifically, we aimed to test measures of cultural similarity based on food and drink interests in social media to predict international migration flows. Measures of cultural distance are difficult to estimate and thus have not yet been widely adopted in gravity models for assessing and predicting migration. However, culture plays an important role in the processes of migration.

As was mentioned before, the relationship between migration and culture is likely bidirectional, since cultural fit in terms of language, norms, and values is an important factor that people consider before moving between countries, and migrants transmit cultural elements from their origin country to their home country and back again during the migration process. In this paper, we focused on showing how measures of cultural similarity derived from Facebook users' food and drink interests can be used to explain migration flows between countries. For instance, imagine that the number of Facebook users living in the United States who are interested in some traditional dishes from Brazil has increased. One possible reason for this development is that the number of Brazilian immigrants in the United States has increased, and thus the number of Americans who are exposed to Brazilian interests has risen. In this example, if these Brazilian immigrants established a big Brazilian community in the United States, the number of Brazilian immigrants could increase even more. In this case, the number of Facebook users interested in Brazilian food and drink serves as a proxy for the Brazilian community established in the United States. One of our main results shows the importance of the cultural similarity between countries, as measured by Facebook users' interests in food and drink, for predicting migration flows between these countries.[20] While our study has broader ramifications, the scope of this article is more limited, as we showed the positive association between cultural similarity and migration flows without attempting to establish a causal direction in this complex bidirectional relationship. Future studies could collect additional data and develop new methods to address this issue and move toward providing more causal estimates of the direction of the relationship between culture and migration flows.

Caution should be exercised when interpreting our results due to their limitations, which we would like to acknowledge. First, the present analysis is constrained by data availability: only 16 countries were included in our analysis. The 16 countries selected for the analysis were chosen to match the list of countries included in Silva et al. (2014) in order to enable us to compare the results from different types of social media data. As well as to ensure the comparability of our findings with those of other studies, we selected these 16 countries in order to cover a large and diverse portion of the world's regions and to include countries where the Facebook penetration rate is high, thus reducing the potential size of the biases in the data. The

data collection could be extended to more countries. However, while the Facebook audiences' interests for the most current period could be collected, the lack of adequate migration data remains a crucial bottleneck. The last time period for which global estimations of migration flow data are available from our main source (Abel and Cohen 2019) is 2015–2019. In other words, we do not have migration flow data, or even migration stock data, after 2019. Moreover, the COVID-19 pandemic affected migration, and we do not have updated data that we could use as a dependent variable in our models. Once new migration data are available, new data from Facebook can be collected in real-time, and the predictions can be updated.

The measures of similarity that we used relied only on data regarding Facebook users' interests in food and drink. Although the cuisine of a country is an important cultural marker for studying cultural similarity, the proposed methodology could be used with other types of attributes and interests, which might be relevant for studies with other goals or angles. We expect that a broader operationalization of measures of culture would lead to the development of models with even higher predictive accuracy. In this sense, what we showed is likely a lower bound in terms of predictive capacity.

Moreover, the social media data we used, including the Facebook Ads data, and the interest categories provided by Facebook may not be exhaustive or representative. Facebook data include a number of biases, given that Facebook users are not necessarily representative of the underlying population in their respective countries. There is a growing literature that has expanded our knowledge on how to identify and correct biases in social media data.[21] In addition to representativity, the classification of users into the categories provided on Facebook Ads could be a source of bias. Grow et al. (2022) evaluated the bias regarding location, age, and gender on Facebook. The authors compared the information provided by the participants of an anonymous online survey with Facebook Ads' classification of the same individuals. The results showed that about 86–93% of respondents' answers matched Facebook's classification. Although location, age, and gender appear to be identified mostly correctly on Facebook Ads, the accuracy of the classification of Facebook users' interests has not been tested systematically. We hypothesize that Facebook covers only a subset of users' interests, but future research is needed to assess the extent to which the representation of interests is accurate.

We would like to emphasize the importance of addressing issues such as biases in digital trace data, and we point the readers to the resources mentioned above for a series of approaches developed to tackle this problem. With our article, we are entering partially uncharted territory in terms of assessing the biases related to our methods, as our approaches are novel, and are not yet part of the conventional toolbox. We hope that our study will further stimulate methodological research on identifying and correcting

biases when studying cultural dimensions using social media data. While the reader should be aware that the data used in this article are not necessarily representative of the entire underlying populations, we should also note that our decision to focus on 16 countries with high Facebook penetration rates limited the extent of the biases, as it allowed for comparisons across countries where Facebook is used in relatively similar ways by comparable demographic segments of the population. It is noteworthy that, despite the biases, the predictive model performs very well. Once the biases are fully modeled, we expect that the predictive capacity will increase. We hope that this article lays the foundation for further analyses that can help us better understand these data and their potential, especially in countries and contexts that have historically been data-poor.

Conclusion

In this paper, we demonstrated that measures of cultural similarity derived from survey and social media data can be important variables in predictions of migration flows. We compared a measure of cultural similarity derived from the WVS with measures of food and drink similarity derived from Foursquare and Facebook Ads data. By using the measures derived from the Facebook Ads data, we introduced a more nuanced view of symmetric and asymmetric measures of similarity and showed how these measures of similarity can be used to explain migration flows between countries. Our results indicated that the Facebook measures of food and drink similarity can play an important role in predicting migration, as they are comparable to standard predictors, such as shared language and shared history. Finally, while we found that some variables, such as shared language, history, and geographic distance, are static and symmetric, we also observed that cultural attributes from daily life are sensitive to changes in the environment and can be represented as an asymmetric measure of similarity between countries, thus adding value to models of migration from both a substantive and a predictive perspective.

Acknowledgments

The authors gratefully acknowledge the resources provided by the International Max Planck Research School for Population, Health, and Data Science (IMPRS-PHDS) and the Max Planck Institute for Demographic Research (MPIDR).

Data availability statement

According to Facebook's Terms of Service, the raw data collected from the Facebook Advertising Platform cannot be shared publicly. In this case, we

do not share the raw data. Instead, the repository contains all the data used in our models, including the measures derived from the Facebook Ads data and the code to replicate all the analyses and generate the figures (see https://github.com/carolcoimbra/gravity-fb).

Notes

1 Although Facebook provides a range of interests broadly related to food and drink, most of those interests do not represent food as naturally found in nature (e.g., grapes, corn). Additionally, Vieira et al. (2022) manually validated the data set by removing interests such as restaurants and brand names. The majority of the interests related to food and drink on Facebook represent dishes or any food or drink processed by humans (e.g., wine, quesadilla).

2 https://ich.unesco.org/en/lists?term[]=vocabulary_thesaurus-10

3 https://www.hofstede-insights.com/models/national-culture

4 https://ourworldindata.org/explorers/migration?time=latest&facet=none&Metric=Net+migration+rate&Period=Total&Sub-metric=Total

5 https://journals.plos.org/plosone/article?id https://doi.org/10.1371/journal.pone.0262947

6 https://github.com/carolcoimbra/cultural-similarity-fb

7 We conducted additional analyses to show how stable the results are when we vary the number of interests we consider in the Facebook measures of similarity. The top 50 foods and drinks generate the best results based on all the calculated metrics, such as the adjusted R-squared, and significant coefficients.

8 https://brandongaille.com/26-great-foursquare-demographics

9 https://www.statista.com/statistics/814726/share-of-us-internet-users-who-use-foursquare-by-age

10 https://99firms.com/blog/foursquare-statistics/#gref

11 https://financesonline.com/foursquare-statistics

12 https://foursquare.com/products/pricing

13 https://www.worldvaluessurvey.org/WVSContents.jsp

14 The most recent seventh wave of the WVS (2017-2022) covers 80 countries.

15 https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp

16 https://www.un.org/development/desa/pd/content/international-migrant-stock

17 https://databank.worldbank.org/home

18 The logarithm scale used in this study is the logarithm base 10.

19 https://www.facebook.com/iq/insights-to-go/2740m-facebook-monthly-active-users-were-2740m-as-of-september-30

20 We conducted additional analyses to investigate the role of the immigrant community in the host country in shaping the significant coefficients observed for cultural similarities. We added migration stocks from 2019 to all the models and observed that overall the coefficients regarding cultural similarities decreased by 30% but were still significant. This result indicates that even though food and drink from the origin country may have been introduced to the destination country by immigrants, the interest in those food and drink cannot be fully explained by the size of the immigrant population. In other words, the interest in food and drink from the origin country is spread across the population in the destination country.

21 There is a growing literature that has expanded our knowledge on how to identify and correct biases in social media data. One line of research has focused on identifying the different types of errors and biases in studies that use digital trace data and on or-

ganizing them in a framework (Olteanu et al. 2019; Sen et al. 2021). For instance, Sen et al. (2021) proposed a categorization based on the total survey error framework to identify several types of errors that may occur in studies that use digital traces. As a consequence, these frameworks also contribute to creating a common vocabulary between researchers using digital trace data. In addition, Drouhot et al. (2023) provided an overview of how some innovative data sets and methodological tools can enrich migration research. Despite all the advantages and promises of using digital trace data for migration research (e.g., less time and costs needed to leverage data for a large sample size), the authors pointed out some of the challenges that can arise when working with these data. Since digital trace data are not generated for research purposes, some extra care is required to repurpose their meaning, as they might otherwise be too superficial or inappropriate for addressing many central research questions. Besides concerns about data quality, some of the key challenges involved in working with these data are related to ethical considerations and selection bias.Zagheni and Weber (2015) considered the problem of selection bias in nonrepresentative samples, such as digital trace data, and proposed two main approaches to reduce bias: the calibration approach and the difference-in-differences approach. In the calibration approach, the online data are adjusted based on reliable official statistics, including through the generation of correction factors (Zagheni and Weber 2012; Zagheni, Weber, and Gummadi 2017; Ribeiro, Benevenuto, and Zagheni 2020). For instance, Ribeiro, Benevenuto, and Zagheni (2020) compared data from Facebook Ads and the US Census and calculated correction factors for some demographic dimensions, such as age, gender, education, and income. However, for contexts where no reliable statistical data are available, the authors suggested a difference-in-differences approach to evaluate relative changes pre- and post-event (Flores 2017; Alexander, Polimis, and Zagheni 2019). Alexander, Polimis, and Zagheni (2019) used Facebook Ads data and the difference-in-differences approach to monitor flows of outmigrants from Puerto Rico before and after Hurricane Maria in 2017. The difference-in-differences approach assumes a constant relationship between estimates from digital trace data and official data, at least over relatively short periods of time.An emerging line of research focuses on using Bayesian approaches to combine different sources of data to estimate migration trends (Rampazzo et al. 2021; Alexander, Polimis, and Zagheni 2020; Hsiao et al. 2023). For instance, in a recent study focused on nowcasting stocks of migrants in the United States, Alexander, Polimis, and Zagheni (2020) demonstrated that a Bayesian hierarchical model combining data from both Facebook and the American Community Survey outperforms alternative models that use only Facebook data or that solely rely on time-series data from the American Community Survey. Recently, Leasure et al. (2023) built a real-time monitoring system to estimate subnational population sizes and internal displacement in Ukraine by leveraging data from Facebook Ads in combination with pre-conflict population data in Ukraine.

References

Abel, Guy J., and Joel E. Cohen. 2019. "Bilateral International Migration Flow Estimates for 200 Countries." *Scientific Data* 6(1): 1–13.

Abel, Guy J., James Raymer, and Qing Guan. 2019. "Driving Factors of Asian International Migration Flows." *Asian Population Studies* 15(3): 243–265.

Alexander, Monica, Kivan Polimis, and Emilio Zagheni. 2019. "The Impact of Hurricane Maria on Out-Migration from Puerto Rico: Evidence from Facebook Data." *Population and Development Review* 45(3): 617–630.

Alexander, Monica, Kivan Polimis, and Emilio Zagheni. 2020. "Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States." *Population Research and Policy Review* 41: 1–28.

Almerico, Gina M. 2014. "Food and Identity: Food Studies, Cultural, and Personal Identity." *Journal of International Business and Cultural Studies* 8(1): 1–7.

Anderson, James E. 2011. "The Gravity Model." *Annual Review of Economics* 3(1): 133–160.

Ashley, Bob, Joanne Hollows, Steve Jones, and Ben Taylor. 2004. *Food and Cultural Studies*. London: Routledge.

Axelson, M. L. 1986. "The Impact of Culture on Food-Related Behavior." *Annual Review of Nutrition* 6(1): 345–363.

Beine, Michel, Simone Bertoli, and Jesús Fernández-Huertas Moraga. 2016. "A Practitioners' Guide to Gravity Models of International Migration." *The World Economy* 39(4): 496–512.

Beyer, Robert M., Jacob Schewe, and Hermann Lotze-Campen. 2022. "Gravity Models Do Not Explain, and Cannot Predict, International Migration Dynamics." *Humanities and Social Sciences Communications* 9(1): 1–10.

Bijak, Jakub, and Jakub Bijak. 2022. "Uncertainty and Complexity: Towards Model-Based Demography." In *Towards Bayesian Model-Based Demography: Agency, Complexity and Uncertainty in Migration Studies*, pp. 13–29. Cham: Springer.

Böhme, Marcus H., André Gröger, and Tobias Stöhr. 2020. "Searching for a Better Life: Predicting International Migration with Online Search Keywords." *Journal of Development Economics* 142: 102347.

Caragliu, Andrea, Chiara Del Bo, Henri LF de Groot, and Gert-Jan M. Linders. 2013. "Cultural Determinants of Migration." *The Annals of Regional Science* 51(1): 7–32.

Cohen, Joel E., 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): 15269–15274.

Counihan, Carole, and Penny Van Esterik, eds. 2012. *Food and Culture: A Reader*. London: Routledge.

De Santis, Gustavo, Mauro Maltagliati, and Silvana Salvini. 2016. "A Measure of the Cultural Distance between Countries." *Social Indicators Research* 126(3): 1065–1087.

De Solier, Isabelle, and Jean Duruz. 2013. "Food Cultures: Introduction." *Cultural Studies Review* 19(1): 4–8.

Drouhot, Lucas G., Emanuel Deutschmann, Carolina V. Zuccotti, and Emilio Zagheni. 2023. "Computational Approaches to Migration and Integration Research: Promises and Challenges." *Journal of Ethnic and Migration Studies* 49(2): 389–407.

Dubois, Antoine, Emilio Zagheni, Kiran Garimella, and Ingmar Weber. 2018. "Studying Migrant Assimilation through Facebook Interests." In *Social Informatics: Proceedings of the 10th International Conference* (SocInfo 2018), St. Petersburg, Russia, September 25–28, 2018, pp. 51–60. Lecture Notes in Computer Science, Vol. 11186. Cham: Springer International Publishing.

Esses, Victoria M. 2018. "Immigration, Migration, and Culture." In *Oxford Research Encyclopedia of Psychology*. Oxford, UK: Oxford University Press.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8(2): 195–222. https://doi.org/10.1093/acrefore/9780190236557.013.287

Ferguson, Gail M., Maria I. Iturbide, and Marcela Raffaelli. 2020. "Proximal and Remote Acculturation: Adolescents' Perspectives of Biculturalism in Two Contexts." *Journal of Adolescent Research* 35(4): 431–460.

Flores, René D. 2017. "Do Anti-immigrant Laws Shape Public Sentiment? A Study of Arizona's SB 1070 Using Twitter Data." *American Journal of Sociology* 123(2): 333–384.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 1995. *Bayesian Data Analysis*. London: Chapman and Hall/CRC.

Ghemawat, P. 2001. "Distance Still Matters." *Harvard Business Review* 79(8): 137–147.

Groves, Robert M., and Lars Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74(5): 849–879.

Grow, André, Daniela Perrotta, Emanuele Del Fava, Jorge Cimentada, Francesco Rampazzo, Sofia Gil-Clavel, Emilio Zagheni, René D. Flores, Ilana Ventura, and Ingmar Weber. 2022. "Is Facebook's Advertising Data Accurate Enough for Use in Social Science Research? Insights from a

Cross-National Online Survey." *Journal of the Royal Statistical Society Series A: Statistics in Society* 185(Supplement 2): S343–S363.

Gupta, Vipin, Paul J. Hanges, and Peter Dorfman. 2002. "Cultural Clusters: Methodology and Findings." *Journal of World Business* 37(1): 11–15.

Hofstede, Geert. 1983. "National Cultures in Four Dimensions: A Research-Based Theory of Cultural Differences among Nations." *International Studies of Management & Organization* 13(1-2): 46–74.

Hsiao, Yuan, Lee Fiorio, Jonathan Wakefield, and Emilio Zagheni. 2023. "Modeling the Bias of Digital Data: An Approach to Combining Digital With Official Statistics to Estimate and Predict Migration Trends." *Sociological Methods & Research*. OnlineFirst. https://doi.org/10.1177/00491241221140144

Inglehart, Ronald. 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton, NJ: Princeton University Press.

Kaasa, Anneli, and Michael Minkov. 2022. "Are Different Two-Dimensional Models of Culture Just a Matter of Different Rotations? Evidence from the Analysis Based on the WVS/EVS." *Journal of Cross-Cultural Psychology* 53(2): 127–156.

Kim, Keuntae, and Joel E. Cohen. 2010. "Determinants of International Migration Flows to and from Industrialized Countries: A Panel Data Approach beyond Gravity." *International Migration Review* 44(4): 899–932.

Kittler, Pamela Goyan, Kathryn P. Sucher, and Marcia Nelms. 2016. *Food and Culture*. Boston, MA: Cengage Learning.

Kosinski, Michal, 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): 543.

Kwantes, Catherine T., Sharon Glazer, Catherine T. Kwantes, and Sharon Glazer. 2017. "Toward an Operationalization of Culture." In *Culture, Organizations, and Work: Clarifying Concepts*, 13–43. Berlin: Springer.

Laitin, David D. 2000. "What Is a Language Community?" *American Journal of Political Science* 44: 142–155.

Leasure, Douglas R., Ridhi Kashyap, Francesco Rampazzo, Claire A. Dooley, Benjamin Elbers, Maksym Bondarenko, Mark Verhagen. et al. 2023. "Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data." *Population and Development Review* 49: 231–254.

Lewer, Joshua J., and Hendrik Van den Berg. 2008. "A Gravity Model of Immigration." *Economics Letters* 99(1): 164–167.

Massey, Douglas S., 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): 431–466.

Mayer, Thierry, and Soledad Zignago. 2011. "Notes on CEPII's Distances Measures: The GeoDist Database." Working Papers 2011-25, Paris: CEPII.

Mayer, Thierry, and Soledad Zignago. 2005. "Market Access in Global and Regional Trade." Paris: CEPII.

Melitz, Jacques, and Farid Toubal. 2014. "Native Language, Spoken Language, Translation and Trade." *Journal of International Economics* 93(2): 351–363.

Mesoudi, Alex. 2018. "Migration, Acculturation, and the Maintenance of Between-Group Cultural Variation." *PloS ONE* 13(10): 1–23.

Mohr, John W., Christopher A. Bail, Margaret Frye, Jennifer C. Lena, Omar Lizardo, Terence E. McDonnell, Ann Mische, Iddo Tavory, and Frederick F. Wherry. 2019. *Measuring Culture*, New York Chichester, West Sussex: Columbia University Press. https://doi.org/10.7312/mohr18028

Montanari, Massimo. 2006. *Food is Culture*. New York: Columbia University Press.

Mucciardi, Massimo, and Gustavo De Santis. 2017. "Cultural versus Objective Distances: the DBS-EM Approach." *Social Indicators Research* 130(3): 867–882.

Muthukrishna, Michael, Adrian V. Bell, Joseph Henrich, Cameron M. Curtin, Alexander Gedranovich, Jason McInerney, and Braden Thue. 2020. "Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance." *Psychological Science* 31(6): 678–701.

Obradovich, Nick, Ömer Özak, Ignacio Martín, Ignacio Ortuño-Ortín, Edmond Awad, Manuel Cebrián, Rubén Cuevas, Klaus Desmet, Iyad Rahwan, and Ángel Cuevas. 2022. "Expanding the Measurement of Culture with a Sample of Two Billion Humans." *Journal of the Royal Society Interface* 19(190): 20220085.

Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. "Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries." *Frontiers in Big Data* 2: 13.

Opare-Obisaw, Clara, Docea AG Fianu, and Kezia Awadzi. 2000. "Changes in Family Food Habits: The Role of Migration." *Journal of Consumer Studies & Home Economics* 24(3): 145–149.

Palotti, Joao, Natalia Adler, Alfredo Morales-Guzman, Jeffrey Villaveces, Vedran Sekara, Manuel Garcia Herranz, Musa Al-Asad, and Ingmar Weber. 2020. "Monitoring of the Venezuelan Exodus through Facebook's Advertising platform." *PloS ONE* 15(2): e0230455.

Pedersen, Peder J., Mariola Pytlikova, and Nina Smith. 2004. "Selection or Network Effects? Migration Flows into 27 OECD Countries, 1990–2000." IZA Discussion Paper No. 1104. Bonn: IZA.

Ramos, Raul. 2016. "Gravity Models: A Tool for Migration Analysis." *IZA World of Labor*. Bonn: IZA.

Rampazzo, Francesco, Jakub Bijak, Agnese Vitali, Ingmar Weber, and Emilio Zagheni. 2021. "A Framework for Estimating Migrant Stocks Using Digital Traces and Survey Data: An Application in the United Kingdom." *Demography* 58(6): 2193–2218.

Rapoport, Hillel, Sulin Sardoschau, and Arthur Silve. 2020. "Migration and Cultural Change." Working Papers 2020-10, Paris: CEPII.

Ravenstein, E. G. 1889. "The Laws of Migration." *Journal of the Royal Statistical Society* 48: 167–227.

Recchi, Ettore, and Adrian Favell. 2019. *Everyday Europe: Social Transnationalism in an Unsettled Continent*. Bristol: Policy Press.

Ribeiro, Filipe N., Fabrício Benevenuto, and Emilio Zagheni. 2020. "How Biased Is the Population of Facebook Users? Comparing the Demographics of Facebook Users with Census Data to Generate Correction Factors." In *Proceedings of the 12th ACM Conference on Web Science*, pp. 325–334. New York: ACM.

Schwartz, Shalom H. 1994. "Beyond Individualism/Collectivism: New Cultural Dimensions of Values." In *Individualism and Collectivism: Theory, Method, and Applications*, edited by U. Kim, H. C. Triandis, Ç. Kâğitçibaşi, S.-C. Choi, and G. Yoon, pp. 85–119. Thousand Oaks, CA: Sage.

Sen, Indira, Fabian Flöck, Katrin Weller, Bernd Weiß, and Claudia Wagner. 2021. "A Total Error Framework for Digital Traces of Human Behavior on Online Platforms." *Public Opinion Quarterly* 85(S1): 399–422.

Silva, Thiago, Pedro Vaz De Melo, Jussara Almeida, Mirco Musolesi, and Antonio Loureiro. 2014. "You Are What You Eat (and Drink): Identifying Cultural Boundaries by Analyzing Food and Drink Habits in Foursquare." *Proceedings of the International AAAI Conference on Web and Social Media* 8(1): 466–475.

Speicher, Till, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. "Potential for Discrimination in Online Targeted Advertising." Paper presented at Conference on Fairness, Accountability and Transparency. *Proceedings of Machine Learning Research* 81: 1–15.

Spyratos, Spyridon, Michele Vespe, Fabrizio Natale, Ingmar Weber, Emilio Zagheni, and Marzia Rango. 2019. "Quantifying International Human Mobility Patterns Using Facebook Network Data." *PloS ONE* 14(10): e0224134.

Suchman, Edward A. 1962. "An Analysis of "Bias" in Survey Research." *Public Opinion Quarterly* 26: 102–111.

Taras, Vas, Julie Rowney, and Piers Steel. 2009. "Half a Century of Measuring Culture: Review of Approaches, Challenges, and Limitations Based on the Analysis of 121 Instruments for Quantifying Culture." *Journal of International Management* 15(4): 357–373.

Tinbergen, Jan. 1962. *Shaping the World Economy; Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund.

Tung, Rosalie L., and Alain Verbeke. 2010. "Beyond Hofstede and GLOBE: Improving the Quality of Cross-cultural Research." *Journal of International Business Studies* 41, 1259–1274.

Vieira, Carolina C., Sophie Lohmann, Emilio Zagheni, Pedro OS Vaz de Melo, Fabrício Benevenuto, and Filipe N. Ribeiro. 2022. "The Interplay of Migration and Cultural Similarity between Countries: Evidence from Facebook Data on Food and Drink Interests." *PloS ONE* 17(2): e0262947.

Vieira, Carolina, Filipe Ribeiro, Pedro Olmo Vaz de Melo, Fabricio 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, pp. 3091–3097. New York: ACM.

White, Roger. 2013. "Is Cultural Distance a Determinant of International Migration Flows? Evidence from Denmark, Germany, and the Netherlands." *Economics Bulletin* 33(3): 2156–2168.

You, Quanzeng, Darío García-García, Mahohar Paluri, Jiebo Luo, and Jungseock Joo. 2017. "Cultural Diffusion and Trends in Facebook Photographs." In *Proceedings of the International AAAI Conference on Web and Social Media* 11(1): 347–356.

Zagheni, Emilio, and Ingmar Weber. 2012. "You Are Where You E-mail: Using E-mail Data to Estimate International Migration Rates." In *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 348–351. New York: ACM.

Zagheni, Emilio, and Ingmar Weber. 2015. "Demographic Research with Non-representative Internet Data." *International Journal of Manpower* 36(1): 13–25.

Zagheni, Emilio, Ingmar Weber, and Krishna Gummadi. 2017. "Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants." *Population and Development Review* 43(4): 721–734.

Zipf, George Kingsley. 1946. "The P 1 P 2/D Hypothesis: On the Intercity Movement of Persons." *American Sociological Review* 11(6): 677–686.