

Digital Diasporas: How Origin Characteristics and Host–Native Distance Shape Immigrants’ Online Cultural Retention

Aparup Khatua¹, David Jurgens², Ingmar Weber³

¹Luleå University of Technology, Sweden

²University of Michigan, Ann Arbor, USA

³Saarland Informatics Campus, Saarland University, Germany
aparup.khatua@ltu.se, jurgens@umich.edu, iweber@cs.uni-saarland.de

Abstract

Immigrants bring unique cultural backgrounds to their host countries. Subsequent interplay of cultures can lead to either a *melting pot*, where immigrants adopt the dominant culture of the host country, or a *mosaic*, where distinct cultural identities coexist. The existing literature primarily focuses on the acculturation of immigrants, specifically the melting pot hypothesis. In contrast, we attempt to identify the antecedents of the mosaic hypothesis or factors that enhance (or diminish) the propensity for cultural retention among immigrants. Based on Facebook advertising data for immigrants from 8 countries residing in the USA, our findings suggest that greater host–native distance is linked to higher online cultural retention, and while origin country context is statistically significant, its impact is generally smaller.

The Interplay of Cultures

If a Colombian immigrant lives in Canada, they will likely be *confronted with the ever-pressing question, “Who am I?” but also, “Am I Canadian?”* Eventually, they may become acculturated into the new culture by embracing social activities of the host nation, such as *playing hockey with Canadian friends, a sport ignored when they lived in Colombia* (Cárdenas and de la Sablonnière 2017, p. 855). This acculturation process helps immigrants develop an identity that enables them to integrate into the unfamiliar terrain of the host country. However, when immigrants move from one country to another, they also carry along their native cultural backgrounds, characterized by unique languages, religions, and societal values (Angelini, Casi, and Corazzini 2015; Cárdenas and de la Sablonnière 2017). The consequence can be either a *melting pot or a mosaic*. In a melting pot scenario, immigrants adopt the dominant culture of the host country, which leads to the homogenization of cultural identities. Conversely, some migrants, regardless of the extent to which they embrace the host culture, maintain their distinct cultural identities, allowing both cultures to coexist like a mosaic that can (potentially) enrich the host country by fostering cultural diversity.

Extant literature on the interplay of cultures has primarily taken a host-country-centric perspective, explored how migrants adapt to their new communities, and assumed that

successful integration primarily depends on the efforts of immigrants (Stewart et al. 2019; Montreuil and Bourhis 2004; Van Oudenhoven and Ward 2013). The underlying assumption of the *melting pot hypothesis* is that over time, this acculturation process would lead to a homogeneous society where the cultural differences would be less pronounced, and our proverbial Colombian immigrant would start playing hockey and consider himself a Canadian. Conversely, according to the *mosaic hypothesis*, immigrants would prefer to maintain their distinct cultural identities. Irrespective of whether they are embracing the host culture, they preserve the traditions and practices of their native country and can contribute to the cultural mosaic of the host country. Ideally, this cultural pluralism will enrich the host country by fostering cross-cultural understanding and appreciation. However, the counter-argument is that a cultural mosaic due to immigrant inflows will be a threat to the cultural identity of the host country (Khatua and Nejd 2023).

Cultural retention among immigrants can potentially foster a sense of belonging and resilience, as well as help immigrants preserve their identity and promote psychological well-being in the host country (Heinonen, Harvey, and Fox 2005; Lillekroken et al. 2024). While offline integration may depend on the host-country context, such as the presence of diaspora communities, online platforms provide spaces for maintaining connections to home cultures. Therefore, in today’s digital era, staying connected to one’s cultural roots increasingly occurs online. For instance, while immigrants might face challenges in maintaining their cultural practices in offline settings, online platforms often provide an alternative space for cultural interactions, such as following media from their home country. However, to the best of our knowledge, no prior studies have investigated the determinants of this form of online cultural retention. Thus, our *Research Questions* are as follows,

RQ1: *Does online cultural retention solely depend on the characteristics of their country of origin?*

RQ2: *Is it also affected by the various dimensions of the distance between the host and native countries?*

Our paper aims to address the above RQs by probing online Facebook data, widely used by prior studies to investigate various cultural aspects of immigrants (Stewart et al. 2019; Khatua, Zagheni, and Weber 2023; Vieira et al. 2020; Dubois et al. 2018). We consider the Facebook Advertising

API to probe the cultural retention of immigrants from eight countries settled in the USA. Facebook API data is not only an innovative proxy for cultural retention but also captures a different aspect of cultural retention. For instance, an immigrant may be outwardly integrated into their offline life but still only listen to music and read news from their home country (Barnett and Benefield 2017; Günsoy et al. 2020). In this way, this study focuses more on online cultural retention than on conventional offline cultural retention.

Following prior studies, we have carefully identified a plausible set of antecedents: both country-specific variables (e.g., economic development, political liberties, civil rights, etc.) as well as different dimensions of distance (e.g., geographic, economic, and Hofstede's 6 Dimensions of Cultural differences) between the host and native countries. To capture cultural retention, we have considered the culinary, musical, and recreational preferences of immigrants (Dubois et al. 2018; Stewart et al. 2019; Vieira et al. 2020, 2022; Obradovich et al. 2022). Our findings suggest that country-specific factors can have mixed effects on cultural retention, but a greater distance between host and native countries enhances cultural retention. Overall, our findings contribute to a better understanding of the multifaceted factors shaping culture retention among immigrants, offering valuable insights to understand cultural diversity within host countries.

Melting Pot or Mosaic?

The metaphor of the melting pot emphasizes assimilation: immigrants are expected to embrace the host nation's norms and values, often linked to higher well-being and social integration. A plethora of studies probed the acculturation process of immigrants in the host country and the consequences of the same. For instance, the literature suggests a positive association between cultural assimilation (of immigrants with the host culture) and immigrants' life satisfaction or subjective well-being, i.e., a strong host identity is associated with high life satisfaction, but this is not the case for native identity (Angelini, Casi, and Corazzini 2015). A weak sense of native identity is associated with higher levels of life satisfaction. Interestingly, these results remain consistent even after controlling for individual characteristics like employment status and income, and notably, "potential benefits of cultural assimilation go beyond the time dimension of the integration process . . . its association with life satisfaction is stronger for established immigrants than for recent ones" (Angelini, Casi, and Corazzini 2015, p. 841). Other studies also pointed out that challenges in cultural assimilation or adaptation are often associated with long-term mental health problems like trauma and distress (Bhugra 2005).

Conversely, the mosaic hypothesis highlights cultural pluralism: immigrants preserve distinct practices and traditions that coexist with those of the host society. Extant literature has highlighted that immigrants retain their native culture in terms of leisure practices (Heinonen, Harvey, and Fox 2005), food and dietary preferences (Lillekroken et al. 2024). This pluralism can enrich the host country by fostering mutual understanding. For instance, the US president has been wishing and celebrating 'Happy Diwali', a festival of

light primarily celebrated in India (Biden 2022). This greeting conveys respect and cultural recognition of the Indian American community settled in the USA. These practices lend support to the mosaic view.

Overall, this debate connects to broader theories of acculturation, such as the seminal work of Berry (1997), which distinguishes between *assimilation* (when individuals adopt the host culture over their own), *separation* (when they reject the host culture to preserve their origin culture), *integration* (when they embrace aspects of both cultures), and *marginalization* (when they disengage from both cultures). The Melting Pot Hypothesis closely aligns with Berry's assimilation, whereas the Mosaic Hypothesis emphasizes pluralism and multiculturalism traditions, i.e., Berry's integration categorization, emphasizing the coexistence of multiple cultural identities within a host society. However, it is also possible that an immigrant may reject the host culture to preserve their native culture, i.e., separation. Broadly, while the melting pot assumes eventual convergence into a dominant culture, the mosaic points out that immigrants may add to the cultural diversity of the host nation.

Notably, these acculturation theories were developed primarily in offline contexts. We argue that cultural retention on online platforms may manifest differently. As noted, immigrants may continue their interests in home-country music, cuisine, or media, even as they participate in offline host-country cultural practices and social activities. By examining how origin-country characteristics and host-native distances shape such online expressions of culture, our study enriches the theoretical discourse of acculturation by providing one of the first large-scale empirical studies on online cultural retention.

Migration Studies using Digital Data

In recent times, extant literature has used digital data to study the cultural dynamics of immigrants. Specifically, Facebook data was considered as a proxy for immigrants' cultural interests (Stewart et al. 2019; Dubois et al. 2018; Vieira et al. 2020). This stream of studies mostly examined the cultural assimilation of migrants, i.e., the melting pot hypothesis. For instance, Stewart et al. (2019) argued that the alignment of musical tastes with host-country preferences indicates higher assimilation and found that Mexican and Indian immigrants in the U.S. exhibited relatively higher cultural convergence than immigrants from East Asia. Similarly, Dubois et al. (2018) highlighted the role of language and origin-country characteristics in shaping assimilation patterns among Arabic-speaking migrants in Germany.

Beyond cultural aspects, extant literature has primarily focused on opinion mining and digital census applications. Twitter-based studies, presently X platform, reveal polarized sentiments toward immigrants: some of the prior studies emphasized a humanitarian perspective, such as struggles of migrants (Siapera et al. 2018; Khatua and Nejd 2022), while other studies portrayed migrants as cultural or security threats (Kreis 2017; Khatua and Nejd 2023). Another stream of research considered online data for performing a digital census, such as estimating migration stocks and flows

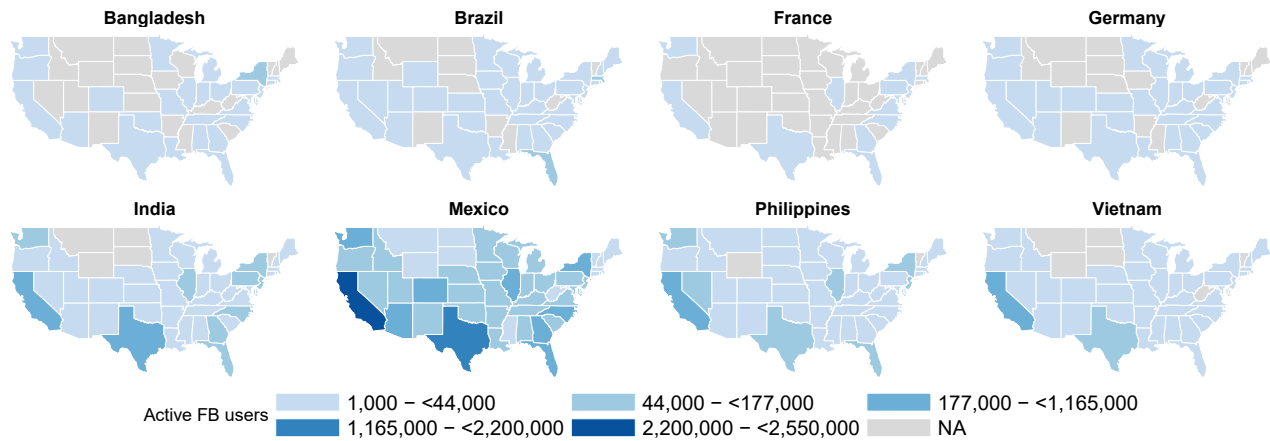


Figure 1: Spatial distribution of *Active Facebook Users* across U.S. states among 8 diaspora groups (expats), according to Facebook categorization. Values are classified using Fisher–Jenks natural breaks, grouping states into distinct clusters based on data distribution. Darker shades represent higher values; grey states indicate data unavailability or sparsity.

using Facebook advertising data (Zaghenei, Weber, and Gum-madi 2017; Rampazzo et al. 2021) or geo-tagged Twitter traces (Zaghenei et al. 2014; Kim et al. 2020). These studies confirm the viability of digital data as an alternative to surveys and censuses. A handful of qualitative studies elucidate that online platforms also provide emotional support and continuity for immigrants, offering ways to sustain connections with families and home cultures (Francisco 2015; Lingel, Naaman, and Boyd 2014; Kaufmann 2018; Khatua 2025).

Research Gap: Extant literature, using digital data, has primarily emphasized assimilation, reinforcing the melting pot perspective. Other studies also point out that social media enables migrants to maintain cultural ties, implicitly supporting the mosaic hypothesis. However, to the best of our knowledge, *no study has systematically examined the conditions under which online cultural retention emerges*, or the role of origin-country characteristics and host-native distances in shaping it. Our study addresses this gap by analyzing Facebook advertising data on the cultural interests of immigrants from 8 countries in the USA. We move beyond cultural assimilation to identify factors that foster or constrain cultural retention, *to investigate the mosaic hypothesis in a digital context*.

Methodology

Country Selection: We selected a total of 8 countries from Asia, Southeast Asia, Latin America, and Europe, two from each region, based on the higher presence of immigrants in the USA. We had to consider the availability of data. For instance, the highest numbers of immigrants in the United States are from Mexico, India, and China¹. However, due to the scarcity of Facebook data in China, we could not con-

sider it in our analysis. Figure 1 shows the state-wise distribution of immigrants, and given the data sparsity even for the 8 countries included, adding more countries would contribute little additional value. Thus, while the list is not exhaustive, it is appropriate to address our research questions.

Facebook Advertising API: Facebook users leave traces of their preferences across diverse interests such as *music, food, and recreation* through interactions like watching videos or commenting on posts. For a specific interest, Facebook provides advertisers an estimated audience size, i.e., the total number of active Facebook users for specific demographics (e.g., age, gender, or education). While it’s not explicitly outlined how these interests are generated or maintained by Facebook, previous research indicates that these interests are dynamic and subject to change over time (Stewart et al. 2019; Dubois et al. 2018). While some interests are shared across most countries, others are highly country-specific (Vieira et al. 2020, 2022). Overall, Facebook allows the extraction of data for a diverse range of interests tailored to different countries (Obradovich et al. 2022).

Accordingly, our focus primarily revolves around country-specific interests; specifically, we consider a wide range of *country-specific cultural interests* - as listed in Table 1. The patterns of interests significantly vary across countries (Vieira et al. 2020; Obradovich et al. 2022); for instance, India boasts a diverse range of cuisine-related interests based on regional distinctions, whereas Germany lacks this diversity. We do agree that it may not be feasible to capture an entire country’s cultural orientation through a few interests, but this limitation lies beyond our control and is determined by Facebook’s algorithms, which aim to reach and engage the maximum number of users. Thus, despite these limitations, it is also worth acknowledging that Facebook’s enormous revenue and profit stem from selling this data for marketing purposes. This implicitly affirms the validity and reliability of these interests.

¹Kramer, S., & Passel, J. S. (2025, August 21). Key findings about U.S. immigrants. Pew Research Center.

Country	Country-specific Cultural Interests
Bangladesh 	Bangladeshi rock, Cinema of Bangladesh, Bangladesh national cricket team, Bengali cuisine, RTV (Bangladesh), Bangladesh Premier League, Bangladesh Football Federation, Nemesis (Bangladeshi band), Black (Bangladeshi band)
Brazil 	Music of Brazil, Brazilian rock, Bossa nova, Brazilian cuisine, Texas de Brazil, Churrascaria, Brazil national football team, Brazil national under-20 football team, Brazilian jiu-jitsu, Campeonato Brasileiro Série A, Campeonato Brasileiro Série B
France 	Culture of France, French cuisine, French wine, Baguette, French hip hop, French pop music, Hellfest (French music festival), France Télévisions, Cinema of France, France national football team, France Football, France national rugby union team, Coupe de France
Germany 	Cinema of Germany, German television comedy, The Voice of Germany, German hip hop, German cuisine, Sport1 (Germany), Germany national football team, Germany national under-21 football team, German football league system, Bundesliga, German Grand Prix
India 	Indian rock, Indian pop, Indian folk music, Indian classical music, Music of India, Cinema of India, Indian cuisine, South Indian cuisine, North Indian cuisine, Indian fast food, Indian Premier League, Indian Super League, Indian Cricket League, Indian Cricket
Mexico 	Music of Mexico, Cinema of Mexico, Regional styles of Mexican music, Mexican pop music, Beer in Mexico, Mexican cuisine, Mexico national football team, Liga MX, Mexican League, Mexican Pacific League, Mexican Football Federation
Philippines 	Music of the Philippines, Sugarfree (Filipino band), Pinoy pop, TV5 (Philippines), Filipino Style Recipe, Philippine cuisine, Mang Inasal Philippines, Filipino martial arts, University Athletic Association of the Philippines, Philippines men’s national basketball team
Vietnam 	Vietnamese cuisine, Music of Vietnam, Vietnamese literature, The Voice of Vietnam, MTV (Vietnam), Vietnam national football team, Tet, Vietnam national under-23 football team, Vietnam’s Next Top Model

Table 1: Country-specific interests from Facebook Advertising API identified for capturing Cultural Retention of immigrants.

Additionally, Facebook has the option to specify the geographical location from which we wish to retrieve audience estimates for a specific interest. Also, the Facebook API allows the extraction of information about immigrants (ex-pats). For example, we can extract audience sizes who previously resided in one country (e.g., Mexico) but currently live elsewhere (e.g., the USA). This option allows us to effectively estimate the number of immigrants (i.e., audience size) from India or Mexico now living in the USA (Khatua, Zagheni, and Weber 2023; Dubois et al. 2018; Stewart et al. 2019). When estimating audience size within their home country, we set the location/country accordingly. However, for the host country, in this study, we set it to the USA.

Operationalization of Variables

Cultural Retention: We measure *cultural retention of immigrants* as their continued interest in their home country’s traditions - specifically in the context of culinary choices (Lillekroken et al. 2024; Vieira et al. 2022), musical (Stewart et al. 2019; Ripani 2023), and recreational preferences (Heinonen, Harvey, and Fox 2005; Dubois et al. 2018). The proliferation of various online platforms allows an immigrant to maintain her connection to her roots (Barnett and Benefield 2017; Croucher and Rahmani 2015; Günsoy et al. 2020). Online platforms enable immigrants to watch movies or TV shows from their native countries and exchange opinions (via online platforms) with family and friends residing in their respective native countries. For instance, the cultural retention of French immigrants can be captured through their digital traces (i.e., watching videos or commenting on posts) related to French cultures (i.e., France-specific interests listed in Table 1). If more French immigrants are following or interacting with French-culture-specific content, this

would indicate higher cultural retention by them.

However, absolute values (i.e., the estimated audience size of these French immigrants) would not reveal much insight. Hence, we need to normalize the same with a base value, i.e., French Facebook users in the USA. Next, we also need to consider the culture-related interests of the French native population staying in their home/native country. To estimate audience size, we consider MAU (monthly active users). Facebook API returns two MAU values: lower bound and higher bound. We have taken an average of these two values as audience size (Khatua, Zagheni, and Weber 2023). However, audience size is a function of Facebook users. Hence, a country with higher Facebook penetration will have a higher audience size. To tackle these distortions, we normalize the estimated audience size with total Facebook users. Also, Facebook API may return an audience value less than 1000 - indicating *sparsity*, i.e., Facebook doesn’t have a sufficient number of users for that specific category. To control sparsity, we exclude all data points (equal to or less than 1000) from our final analysis.

Following Dubois et al. (2018) and Stewart et al. (2019), we estimate *Cultural Retention (CR)* for a group of immigrants with respect to natives residing in their home country. Say, we are estimating the cultural orientation of German immigrants in the USA, i.e., our host country ($Host_{Ctry}$). To estimate *Cultural Interests of Immigrants from a specific native country*, i.e., $CI(Imm)_{Ctry}$, we need to estimate the proportion of German immigrants interested in *distinct* Germany or Country-specific Cultural Interests ($Ctry_{CI}$).

$$CI(Imm)_{Ctry} = \frac{\# \text{ of immigrants interested in } Ctry_{CI}}{\# \text{ of immigrants in the } Host_{Ctry}} \quad (1)$$



































Variables	Mean	S.D.	Minimum	Country with Lowest Score	Maximum	Country with Highest Score
Ln(GDP)	14.08	0.983	12.91		15.222	
Ln(GDP/capita)	9.005	1.183	7.788		10.794	
Political Rights Score	26.375	11.856	4.00		39.00	
Civil Liberties Score	35.75	12.948	15.00		54.00	
Linguistic Diversity Index	0.393	0.311	0.100		0.910	
Int'l Tourists Arrival/Population	0.599	1.099	0.002		3.233	
Ln(Immigrants in the USA)	12.33	1.233	10.69		14.208	
Human Development Index	0.769	0.109	0.644		0.95	
Ln(Geographic Distance)	9.079	0.579	7.811		9.526	
Ln(GDP Difference)	16.974	0.062	16.877		17.036	
Ln(GDP/capita Difference)	10.963	0.386	10.226		11.211	
Ln(Power Distance Difference)	4.198	0.235	3.829		4.654	
Ln(Individualism Difference)	4.786	0.235	4.394		5.043	
Ln(MAS Difference)	4.666	0.100	4.533		4.804	
Ln(Uncertainty Avoidance Difference)	4.424	0.244	4.094		4.754	
Ln(Long Term Orientation Difference)	4.659	0.123	4.500		4.844	
Ln(Indulgence Difference)	4.784	0.231	4.263		4.997	

Table 2: Data Overview of Variables

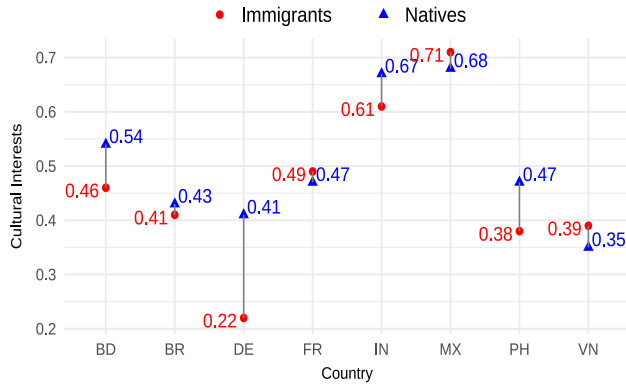


Figure 2: Red dots report the median values of $CI(Imm)_{Ctry}$, i.e., cultural interests across various demographic categories of immigrants residing in the US. Conversely, blue triangles report the median values of $CI(Nat)_{Ctry}$, i.e., similar cultural interests of natives - staying in their respective home countries. This figure indicates wide cross-country variations, e.g., a higher gap for DE and a lower gap for BR and MX.

Next, to estimate *Cultural Interests of Natives of the same country*, i.e., $CI(Nat)_{Ctry}$, we need to estimate the proportion of natives (or Germans residing in Germany) displaying preferences for the *same set of distinct German or Country-specific Cultural Interests* ($Ctry_{CI}$).

$$CI(Nat)_{Ctry} = \frac{\# \text{ of natives interested in } Ctry_{CI}}{\# \text{ of natives in the Home}_{Ctry}} \quad (2)$$

Conceptually, similar values of $CI(Imm)_{Ctry}$ and

$CI(Nat)_{Ctry}$ will indicate no significant differences between immigrants and natives, i.e., broadly similar cultural orientations irrespective of where they are residing. However, Figure 2 indicates variations between natives and immigrants. Subsequently, we also explore the state-wise distribution of cultural interests of immigrants, i.e., $CI(Imm)_{Ctry}$, and report it in Figure 3. With the exception of a few countries, such as Mexico or India, and a few states, such as Texas, data scarcity remains a major concern. This limitation prevents us from conducting a granular sub-national analysis and restricts our analysis to the country level. Following Dubois et al. (2018) and Stewart et al. (2019), we operationalize *Cultural Retention* (CR) as follows:

$$CR_1 = \frac{CI(Imm)_{Ctry}}{CI(Nat)_{Ctry}} \quad (3)$$

For robustness, we also consider CR as follows:

$$CR_2 = Ln \frac{CI(Imm)_{Ctry}}{CI(Nat)_{Ctry}} \quad (4)$$

Native Country Characteristics: We have considered country-level factors that are likely to shape migrants' values, norms, and behavioral patterns, thereby influencing their cultural retention.

GDP and GDP/capita: The economic status or affordability of an immigrant might impact their ability to embrace the host culture. Hence, we consider the Gross Domestic Product (GDP) of a country (in current USD) and GDP per capita (in current USD) to capture this aspect. We extracted data for these variables from the World Bank². To address significant

²<https://data.worldbank.org/indicator>

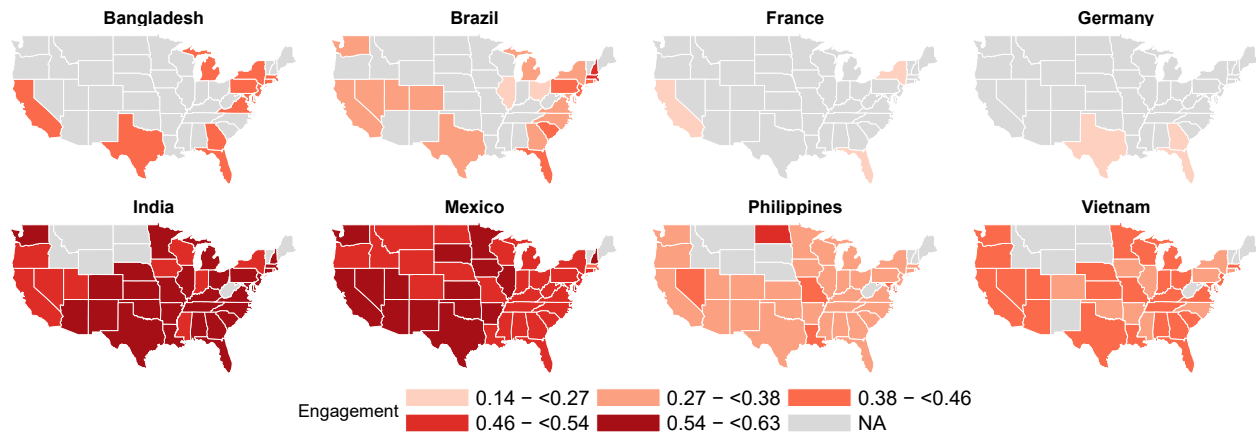


Figure 3: State-level distribution of *own-culture engagement* across 8 diaspora groups. Values correspond to the spatial pattern of red dots in Figure 2, representing state-wise median values of $CI(Imm)_{Ctry}$. Values are classified using the Fisher-Jenks natural breaks method. Grey states denote data sparsity, which is more pronounced than in Figure 1, as that figure reflects the presence of expats (absolute audience size), whereas this figure reports median engagement levels.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	0.97 ***	1.41 ***	1.11 ***	1.19 ***	1.34 ***	1.05 ***	0.97	-0.17	1.12 ***
Age: 35 to 55	0.03	0.04	0.04	0.04	0.05	0.03	0.02	0.04	0.04
Age: 56 and above	0.05	0.05	0.05	0.07	0.08*	0.05	0.02	0.05	0.05
Male Dummy	-0.08*	-0.08*	-0.08*	-0.09**	-0.10**	-0.07*	-0.07*	-0.09**	-0.08*
Edu: Advanced	-0.09*	-0.09*	-0.09*	-0.07	-0.06	-0.09*	-0.10**	-0.08*	-0.09*
Edu: Basic	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.06	-0.05	-0.04
Economic Status: iOS users	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02
Ln(GDP)		-0.03							
Ln(GDP/capita)			-0.02						
Political Rights Score				-0.01 ***					
Civil Liberties Score					-0.01 ***				
Linguistic Diversity Index						-0.18 ***			
Int'l Tourists Arrival/Population							0.03		
Ln(Immigrants in the USA)								0.09 ***	
Human Development Index									-0.21
R^2	0.05	0.07	0.06	0.23	0.28	0.11	0.16	0.16	0.06
Adj. R^2	0.03	0.04	0.02	0.20	0.25	0.08	0.13	0.13	0.03
Max. VIF value	1.01	1.02	1.04	1.02	1.03	1.03	1.01	1.02	1.05
Num. of Obs.	214.00	214.00	214.00	214.00	214.00	214.00	214.00	214.00	214.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3: Reports the OLS regression analysis regarding the Effects of Country Characteristics on Culture Retention of Immigrants. Base categories: Age 18–34, Gender: Female, Edu: Not disclosed, and OS: Android Users. Dependent variable: CR_1 . Robustness tests: Bonferroni correction reconfirms the OLS-based regression analysis at $p < 0.05$ for the variables highlighted in blue.

disparities among countries, we logarithmically transformed these variables for our analysis.

Human Development Index: Unlike the unidimensional GDP/capita, the Human Development Index (HDI) is a composite index that considers three basic dimensions of human development: Health, Education, and Standard of living. Thus, HDI effectively captures the overall well-being and quality of life in the native country. Therefore, in our

analysis, we also consider HDI^3 .

Political Rights and Civil Liberties: Socio-political freedom in the native country can influence behavior and attitude towards diversity or plurality in opinions, which can, in turn, negatively impact the propensity of cultural retention, *refuting the mosaic hypothesis*. Hence, we consider these two aspects in our analysis and extract the data from the Freedom

³<https://hdr.undp.org/data-center/country-insights#/ranks>

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Intercept	0.83 ***	0.75 ***	0.77 ***	0.85 ***	0.71 ***	0.67 ***	0.03	0.77 ***
Age: 35 to 55	0.02	0.02	0.02	0.02	0.01	0.01	0.02	0.02
Age: 56 and above	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.02
Male Dummy	-0.04 *	-0.04 *	-0.05 **	-0.05 **	-0.04 *	-0.04 *	-0.05 **	-0.04 *
Edu: Advanced	-0.04	-0.04	-0.03	-0.03	-0.04 *	-0.04 *	-0.04	-0.04
Edu: Basic	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
Economic Status: iOS users	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Ln(GDP)	-0.01							
Ln(GDP/capita)		-0.01						
Political Rights Score			-0.00 ***					
Civil Liberties Score				-0.01 ***				
Linguistic Diversity Index					-0.09 ***			
Int'l Tourists Arrival/Population						0.01		
Ln(Immigrants in the USA)							0.05 ***	
Human Development Index								-0.14
R^2	0.06	0.05	0.19	0.25	0.10	0.05	0.18	0.06
Adj. R^2	0.02	0.02	0.16	0.22	0.07	0.02	0.15	0.02
Max. VIF value	1.02	1.04	1.02	1.03	1.00	1.02	1.02	1.05
Num. obs.	214.00	214.00	214.00	214.00	214.00	214.00	214.00	214.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4: Robustness Analysis: Effects of Country Characteristics on Culture Retention of Immigrants. Base categories for demographic variables in all regression models are Age: 18–34, Gender: Female, Edu: Not disclosed, and OS: Android Users. Dependent variable: CR₂. **Bonferroni correction** reconfirms the OLS-based regression analysis at $p < 0.05$ for the **variables highlighted in blue**.

House website⁴. Higher scores indicate greater freedom.

Linguistic Diversity: Similar to socio-political freedom, linguistic diversity, i.e., the number of languages spoken within a country, varies significantly across countries. Some nations have only one language, while others have multiple languages or at least different dialects spoken in various regions. We assume exposure to diverse cultures/languages would negatively impact cultural retention, *discarding the mosaic hypothesis*. We consider the Linguistic Diversity Index⁵ for our analysis.

Exposure to international tourism: The opportunity to interact with international tourists in the native country exposes locals (i.e., potential immigrants) to different cultures, fostering greater cultural understanding. Thus, immigrants from countries known for international tourism may have a lower propensity to retain native culture, *contradicting the mosaic hypothesis*. We captured this aspect by dividing international tourism (i.e., number of arrivals) by population. To nullify the COVID-related negative impact on tourism, we used 2019 data points from the World Bank⁶.

Exposure to Compatriots: Conversely, the opportunity to interact with compatriots or co-nationals in the host country allows an immigrant to maintain their own cultural identities. These immigrants may gather or come together during festivals or cultural occasions, *confirming the mosaic hypothesis*. We capture this effect by incorporating

country-wise data points of USA immigrants⁷. To control heteroscedasticity, we performed a log transformation.

Different Dimensions of Distance: For our RQ2, we have considered three dimensions of distance as follows: geographic distance, economic distance, and cultural distance (Coimbra Vieira, Lohmann, and Zagheni 2024).

Geographic Distance: The greater the geographic distance between two locations, the more challenging it can be to travel or communicate. Consequently, less frequent travel to the native country can encourage immigrants to adopt the host culture. Considering the wide disparities in this variable, we have performed log transformations of geographic distances (Mayer and Zignago 2011).

Economic Distance: Previously, we considered the economic status of native countries, but it is worth noting that the disparity between two countries can be a *better predictor of cultural retention or the mosaic hypothesis*. For instance, it would often be challenging for an immigrant from a less economically developed country to afford or participate in the cultural activities of a more affluent country like the USA. Therefore, we considered the difference or gap in GDP and GDP/capita between the USA and the respective native countries and performed a log transformation on these variables for our analysis.

Cultural Distances: Following Tadesse and White (2010, p. 147), we conceptualize cultural distance “as a measure of the degree to which shared norms and values in one country differ from those in another country.” We consider cultural distances across *Hofstede’s 6 Cultural Dimensions* (Hofst-

⁴<https://freedomhouse.org/countries/freedom-world/scores>

⁵<https://worldpopulationreview.com/country-rankings/linguistic-diversity-index-by-country/>

⁶<https://data.worldbank.org/indicator>

⁷<https://worldpopulationreview.com/country-rankings/us-immigration-by-country>

ede and Bond 1984, 1988; Taras, Kirkman, and Steel 2010)

However, is Hofstede's framework, developed in the 1960s and 1970s, still relevant today? Beugelsdijk, Maseland, and Van Hoorn (2015) argue that while cultures have changed slightly, the relative differences between countries remain largely the same, making the framework a reliable measure of cultural distance. We operationalize this variable as follows: Hofstede's Individualism dimension scores for the USA, Brazil, and Vietnam are 60, 36, and 30, respectively. Therefore, we calculate the *Individualism Difference* as $\ln(60-36)$ for Brazil and $\ln(60-30)$ for Vietnam. However, it is worth noting that some countries may have higher scores than the USA on certain cultural dimensions. To avoid negative values, we consider the absolute value before performing logarithmic transformations.

Demographic Characteristics: Following prior studies, we have controlled for different demographic categories (Dubois et al. 2018; Gil-Clavel and Zagheni 2019; Stewart et al. 2019), i.e., we extract audience size for a specific demographic profile. Based on the literature, the following demographic attributes are considered for disaggregation:

Age: 18 to 34; 35 to 55; 56 and above [3 categories]

Gender: Male and Female [2 categories]

Education: Basic (such as High School, Some High School, Undergrad, Associate Degree, etc.); Advanced (such as Professional Degree, Master's Degree, Doctorate Degree, etc.); Undisclosed/unspecified [3 categories]

Economic Status: iOS vs. Android users [2 categories] — prior work uses iOS (Apple) usage as a proxy for higher wealth (Fatehikia et al. 2020, 2022).

Analysis & Findings

Data Snapshot: Table 2 presents summary statistics of explanatory variables, including mean, standard deviation (S.D.), minimum/maximum values, and the corresponding countries with the lowest/highest scores. As anticipated, developed nations, such as Germany, demonstrate robust economic status and socio-political freedom. Additionally, the economic disparities compared to the USA are relatively minimal for Germany. Conversely, significant cultural gaps are observed, particularly between the USA and developing nations such as Vietnam, Mexico, and Bangladesh, where socio-political freedoms are also low.

Statistical Analysis: Tables 3 to 6 report the findings of the Ordinary Least Squares (OLS) regression analysis. For each country, we have extracted audience estimates for 36 combinations: 3 (age) * 2 (gender) * 3 (education) * 2 (economic status). Ideally, we should have 288 observations (36 * 8 countries) for our analysis, but we have only 214 observations due to data sparsity - as evident from Figures 1 and 3. Dependent Variables for Tables 3 and 5 are CR₁.

RQ1: Do Origin-Country Characteristics Matter?

Table 3 reports the impact of various country-specific characteristics on cultural retention among immigrants. Model 1 is our base model with only demographic or control variables. We note that the male gender exhibits a consistent negative association with culture retention across all models. Similarly, advanced education shows a primarily negative

association with culture retention across models, although the significance varies for both variables. Intuitively, male and highly educated immigrants have a higher propensity for social interactions, and this, in turn, negatively impacts their cultural retention tendencies, and *refutes the mosaic hypothesis*. Notably, Model 1 has very low explanatory power (i.e., Adj. R² is 0.03), compared to other models in this table. It strongly indicates that demographic factors (e.g., age, gender, education, or economic wealth) have a limited impact on cultural retention.

Next, we find Political rights, Civil liberties, and Linguistic Diversity emerge as crucial determinants, exhibiting statistically significant ($p < 0.001$) negative associations with culture retention. Intuitively, these findings highlight the importance of political or social freedoms as well as linguistic diversity in the native country in diminishing cultural retention among immigrants, and *our analysis refutes the mosaic hypothesis*. Conversely, factors such as the presence of other immigrants from the native country in the host country demonstrate statistically significant ($p < 0.001$) positive relationships, indicating their impact in fostering cultural retention within immigrant communities, and *supports the mosaic hypothesis*—a vibrant presence of diaspora allows an immigrant to find his own community in the host country and retain his native culture. Surprisingly, we didn't find any statistically significant relationship between the economic health (i.e., GDP and GDP per capita) or HDI of a country and the cultural retention of immigrants. Similarly, we anticipated that exposure to international tourists in the native country would negatively impact cultural retention; however, there was no statistically significant relationship. We speculate that international tourists concentrate on a few destinations or popular cities in a country, and as a result, our variable fails to capture the exposure to diverse cultures. For ease of interpretation, we visually present all model findings together in Figure 5.

Lastly, the Adj. R² values are modest across models, though this is typical for sociocultural studies using social media data. Importantly, our aim is not to develop a predictive model, where Adj. R² would be a primary evaluation metric; rather, our focus is on the direction and significance of theoretically driven relationships.

RQ2: Does Host-Native Distance Matter?

In Table 5, Model 1 incorporates the geographic distance (between the USA and native countries of immigrants), but there was no significant association. Next, we consider economic distance, i.e., differences between the USA and respective native countries in terms of GDP and GDP per capita, and find a statistically significant positive effect. It is worth noting that the direct impact of these variables was not significant in Table 3. Thus, it can be argued that the propensity of cultural retention is high when the gap between the GDP/capita (or affordability) of immigrants and host country citizens is large. Next, we consider cultural distance and incorporate Hofstede's 6 cultural dimensions; however, to avoid multicollinearity issues raised from a VIF analysis, Model 5 considers only Long Term Orientation Difference and reports a statistically significant ($p < 0.05$) positive effect (with the highest VIF value of 1.01), and Model 6 re-

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	1.44 ***	-2.76 **	-10.44	-6.86 **	-0.62	-8.29 ***	0.79 ***
Age: 35 to 55	0.03	0.05	0.04	0.05	0.04	0.06 *	0.05
Age: 56 and above	0.04	0.08	0.06	0.06 *	0.05	0.06 *	0.06 *
Male Dummy	-0.07 *	-0.10 **	-0.08 *	-0.10 ***	-0.08 *	-0.10 ***	-0.10 ***
Edu: Advanced	-0.09 *	-0.07	-0.08 *	-0.07 *	-0.09 *	-0.07 *	-0.07 *
Edu: Basic	-0.05	-0.04	-0.04	-0.06 *	-0.04	-0.06 *	-0.06 *
Economic Status: iOS users	0.02	0.02	0.02	0.03	0.02	0.03	0.03
Ln(Geographic Distance)	-0.05						
Ln(GDP Difference)		0.81 *					
Ln(GDP/capita Difference)			0.33 ***				
Ln(Power Distance Difference)				0.65 ***		0.60 ***	
Ln(Individualism Difference)				0.97 **		0.77 ***	
Ln(Motivation towards Achievement and Success Difference)				1.07 ***		0.97 ***	
Ln(Uncertainty Avoidance Difference)				0.79 ***		0.91 ***	
Ln(Long Term Orientation Difference)				-0.43	0.34 *		
Ln(Indulgence Difference)				-1.29 ***		-1.18 ***	
Country Dummies							Included
R^2	0.07	0.13	0.07	0.57	0.07	0.57	0.57
Adj. R^2	0.04	0.10	0.04	0.55	0.04	0.55	0.54
Max. VIF value	1.01	1.07	1.02	22.55	1.01	4.91	1.14
Num. of Obs.	214.00	214.00	214.00	214.00	214.00	214.00	214.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 5: Effects of Distances on Culture Retention of Immigrants. Base categories: Age 18–34, Female, Edu: Not disclosed, OS: Android Users. Dependent variable: CR₁. Explanatory variables capture differences between the USA and origin countries across geographic, economic, and Hofstede cultural dimensions. Bonferroni correction reconfirms results at $p < 0.05$ for variables highlighted in blue.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	0.96 ***	-5.40	-1.37 **	-3.35 **	-0.33	-4.06 ***	0.57 ***
Age: 35 to 55	0.01	0.02	0.02	0.03	0.02	0.03	0.03
Age: 56 and above	0.02	0.02	0.04	0.03	0.02	0.03	0.03
Male Dummy	-0.04 *	-0.04 *	-0.05 **	-0.05 ***	-0.04 *	-0.05 ***	-0.05 ***
Edu: Advanced	-0.04 *	-0.04	-0.03	-0.03	-0.04	-0.03	-0.03
Edu: Basic	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
Economic Status: iOS users	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Ln(Geographic Distance)	-0.03 *						
Ln(GDP Difference)		0.36 *					
Ln(GDP/capita Difference)			0.18 ***				
Power Distance Difference				0.35 ***		0.32 ***	
Individualism Difference				0.53 ***		0.44 ***	
Motivation towards Achievement and Success Difference				0.50 ***		0.45 ***	
Uncertainty Avoidance Difference				0.40 ***		0.46 ***	
Long Term Orientation Difference				-0.21	0.21 *		
Indulgence Difference				-0.66 ***		-0.61 ***	
Country Dummies							Included
R^2	0.07	0.07	0.13	0.56	0.08	0.55	0.56
Adj. R^2	0.04	0.04	0.10	0.53	0.05	0.53	0.53
Max. VIF value	1.01	1.02	1.07	22.56	1.01	4.91	1.14
Num. obs.	214.00	214.00	214.00	214.00	214.00	214.00	214.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 6: Robustness Analysis: Effects of Distances on Culture Retention of Immigrants. Base categories: Age 18–34, Gender: Female, Edu: Not disclosed, OS: Android Users. Dependent variable: CR₂. Bonferroni correction reconfirms results at $p < 0.05$ for variables highlighted in blue.

ports the other 5 cultural dimensions (with the highest VIF value of 4.91). As anticipated, we find that higher cultural

distance leads to higher cultural retention for 4 dimensions in Model 6 (with $p < 0.001$)—except for Individualism. A

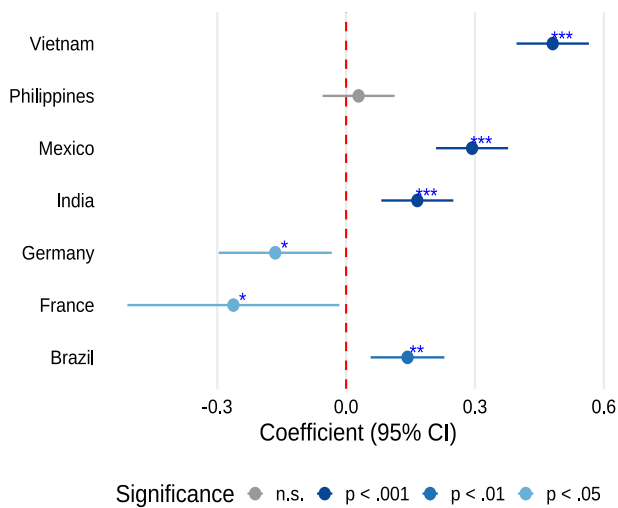


Figure 4: Visual representation of country-level fixed effects on cultural retention (CR_1), with Bangladesh as the base category. The cyan (low statistical significance) and blue (high statistical significance) circles represent the coefficients of country dummy variables estimated in Model M7 (Table 5). Horizontal lines indicate 95% confidence intervals, and blue stars denote significance levels.

negative impact of Individualism indicates immigrants tend to move away from a restrained culture to a more relaxed attitude toward life and embrace the new culture. Notably, $Adj. R^2$ significantly improved from 0.03 (in Model 1 of Table 3) to 0.55 (in Model 6 of Table 5). This improvement in $Adj. R^2$ strongly confirms the impact of cultural distance on cultural retention. Overall, *economic and cultural distances broadly support the mosaic hypothesis*.

It is possible that we may have ignored other determinants of cultural retention, and our list of variables is certainly not exhaustive. Hence, in response to potential omitted variable bias, we have controlled for country effects using dummy variables and found that most country dummies are significant, with a relatively high $Adj. R^2$ of 0.56. For brevity and ease of interpretation, we visually report these country effects in Figure 4.

Visual Representations of OLS Analysis: Figures 4 and 5 visually depict the findings of our OLS analysis. Figure 4 answers whether country matters. From Model 7 of Table 5, the country coefficients and their significance levels are reported in Figure 4, with Bangladesh as the base category. We find that cultural retention scores are significantly lower for Germany and France, two developed countries in our sample, compared to other developing countries, indicating that *the country of origin does matter*. Next, we plot the effects of country characteristics (through red circles) and various dimensions of distance (through cyan triangles) in Figure 5, based on the findings from Tables 3 and 5. Figure 5 strongly suggests that a greater distance is associated with higher cultural retention. For instance, economic distance, captured through GDP difference, has the highest im-

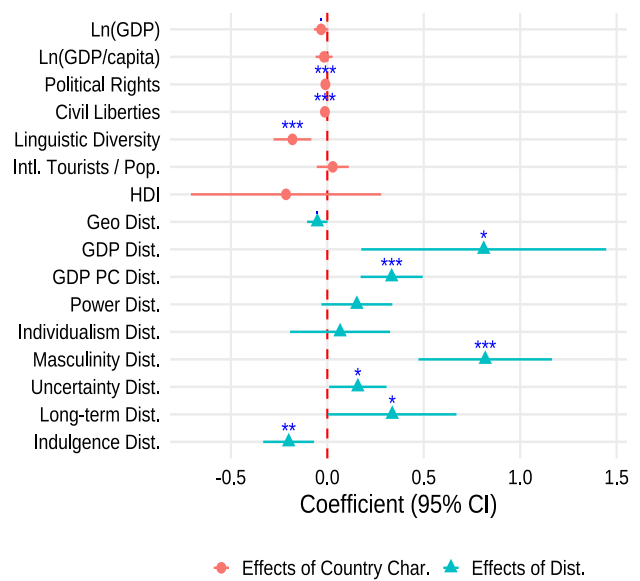


Figure 5: Visual representation of the fixed effects of country characteristics (e.g., GDP, HDI) and distance dimensions (e.g., economic, cultural) on cultural retention (CR_1). The estimated coefficients from regression models in Table 3 are shown as red circles (country characteristics), and those from Table 5 are shown as cyan triangles (distance dimensions). Horizontal lines indicate 95% confidence intervals, and stars denote conventional significance levels ($*p < 0.05$; $**p < 0.01$; $***p < 0.001$). Estimates are standardized for comparability across models.

impact on cultural retention. While country context matters and most results are statistically significant, their impacts are relatively small compared to distance.

Robustness Analysis: For the sake of robustness, we consider an alternate operationalization of Cultural Retention, CR_2 , as the natural logarithm of our original CR_1 .

$$CR_2 = \ln \frac{CI(Imm)_{Ctry}}{CI(Nat)_{Ctry}}$$

In Tables 4 and 6, we repeat our OLS regression analysis with CR_2 as the dependent variable. It is reassuring to note that our results remain consistent for all variables except Geographic Distance. Previously, we didn't find a statistically significant result for $\ln(\text{Geographic Distance})$, but in our robustness analysis, we find a negative and weakly significant effect on cultural retention, *refuting the mosaic hypothesis* - as anticipated. The **Bonferroni correction** is a statistical method used to reduce the risk of falsely identifying a result as significant when performing multiple tests on the same dataset. Our Bonferroni correction reaffirms the results for most key variables - these variables are color-coded in blue in Tables 3, 5, 4 and 6.

Limitations and Future Scope

This study has a few limitations that offer potential avenues for future research. *First*, our analysis is based on cross-

sectional data, and couldn't consider temporal dynamics due to the inherent limitations of Facebook data. However, cultural retention might be stronger during the initial years of migration but could diminish as immigrants integrate into the host society. *Second*, cross-lingual and cross-country comparisons of Facebook interest data have some inherent limitations. For instance, the number of interests available on the platform might vary by country, making normalization of interests across countries/cultures difficult, as reported in Table 1. *Third*, audience sizes for specific interests may fluctuate over time, raising concerns about the stability of findings. *Fourth*, the lack of fine-grained sub-national data, as reported in Figures 1 and 3, limits the study's ability to capture local cultural integration vis-à-vis retention dynamics. Cultural assimilation or retention processes can be influenced by factors such as immigrant density or political leanings (e.g., Republican vs. Democratic regions). *Lastly*, the evidence from low cultural retention provides unclear support for the melting pot hypothesis. According to Berry (1997), low cultural retention might not always lead to separation, rejecting the host culture to retain one's own, but could instead result in marginalization, disengaging from both cultures. Conversely, high cultural retention might not imply assimilation, embracing the host culture over the native one, but could reflect integration, embracing both cultures. The scope of this paper did not allow us to examine these finer nuances, but future studies can explicitly investigate these aspects.

In short, while our study provides valuable insights, such as elucidating the potential of digital data to understand intricate issues like online cultural retention of immigrants, at the same time, the nature of the data restricts us from performing a fine-grained analysis. This implicitly indicates the relevance of a mixed-method collaborative approach, where information science researchers should collaborate with qualitative or survey-based researchers to get a holistic understanding of complex societal issues.

Conclusion

As immigrants navigate the unfamiliar terrain of host countries, they often struggle with their native identity. In this context, the issue of retaining one's native identity (or cultural interests of one's native country) is multifaceted and influenced by a variety of factors. We have explored country-specific characteristics and the distances, geographic, economic, and cultural, between native and host countries to investigate their impact on online cultural retention among immigrants. Our study highlights the significance of both sets of factors, with country-specific variables such as political rights, civil liberties, and linguistic diversity displaying a negative effect on cultural retention. Conversely, the distances between countries, particularly economic and cultural differences, positively impact the cultural retention of immigrants, *aligning with the mosaic hypothesis*. Our study discards a one-size-fits-all approach because integration outcomes depend not only on immigrants' choices but also on structural factors of the host and native countries - a key insight for policymakers striving to support inclusivity within host countries.

Ethical Statement

Facebook's API allows the extraction of anonymous and aggregate-level user data, and it has no re-identification risk for individual users. As we are struggling with issues of integration and cultural identity, this study on cultural retention can potentially enrich our appreciation of diverse cultural practices.

Acknowledgments

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, and by DSO National Laboratories.

References

- Angelini, V.; Casi, L.; and Corazzini, L. 2015. Life satisfaction of immigrants: does cultural assimilation matter? *Journal of Population Economics*, 28: 817–844.
- Barnett, G. A.; and Benefield, G. A. 2017. Predicting international Facebook ties through cultural homophily and other factors. *New Media & Society*, 19(2): 217–239.
- Berry, J. W. 1997. Immigration, acculturation, and adaptation. *Applied psychology*, 46(1): 5–34.
- Beugelsdijk, S.; Maseland, R.; and Van Hoorn, A. 2015. Are scores on Hofstede's dimensions of national culture stable over time? A cohort analysis. *Global Strategy Journal*, 5(3): 223–240.
- Bhugra, D. 2005. Cultural identities and cultural congruency: a new model for evaluating mental distress in immigrants. *Acta Psychiatrica Scandinavica*, 111(2): 84–93.
- Biden, J. 2022. Statement by President Joe Biden on Diwali. White House Press Briefing.
- Cárdenas, D.; and de la Sablonnière, R. 2017. Understanding the relation between participating in the new culture and identification: Two studies with Latin American immigrants. *Journal of Cross-Cultural Psychology*, 48(6): 854–873.
- Coimbra Vieira, C.; Lohmann, S.; and Zagheni, E. 2024. The value of cultural similarity for predicting migration: Evidence from food and drink interests in digital trace data. *Population and Development Review*, 50(1): 149–176.
- Croucher, S. M.; and Rahmani, D. 2015. A longitudinal test of the effects of Facebook on cultural adaptation. *Journal of International and Intercultural Communication*, 8(4): 330–345.
- Dubois, A.; Zagheni, E.; Garimella, K.; and Weber, I. 2018. Studying migrant assimilation through Facebook interests. In *Social Informatics: 10th International Conference, SocInfo 2018, St. Petersburg, Russia, September 25–28, 2018, Proceedings, Part II 10*, 51–60. Springer.
- Fatehkia, M.; del Villar, Z.; Koebe, T.; Letouzé, E.; Lozano, A.; Al Feel, R.; Mrad, F.; and Weber, I. 2022. Using Facebook advertising data to describe the socio-economic situation of Syrian refugees in Lebanon. *Frontiers in Big Data*.
- Fatehkia, M.; Tingzon, I.; Orden, A.; Sy, S.; Sekara, V.; Garcia-Herranz, M.; and Weber, I. 2020. Mapping socio-economic indicators using social media advertising data. *EPJ Data Science*, 9(1): 22.

- Francisco, V. 2015. 'The internet is magic': Technology, intimacy and transnational families. *Critical Sociology*, 41(1): 173–190.
- Gil-Clavel, S.; and Zagheni, E. 2019. Demographic differentials in Facebook usage around the world. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 647–650.
- Günsoy, C.; Olcaysoy Okten, I.; Cross, S. E.; and Saribay, S. A. 2020. Cultural differences in self-expression on Facebook: A Comparison of Facebook Status Updates in Turkey and the USA. *International Journal of Human-Computer Interaction*, 36(18): 1775–1782.
- Heinonen, T.; Harvey, C. D.; and Fox, K. M. 2005. Leisure as part of cultural retention of Finnish-Canadian immigrants. *Leisure/Loisir*, 29(2): 275–300.
- Hofstede, G.; and Bond, M. H. 1984. Hofstede's culture dimensions: An independent validation using Rokeach's value survey. *Journal of cross-cultural psychology*, 15(4): 417–433.
- Hofstede, G.; and Bond, M. H. 1988. The Confucius connection: From cultural roots to economic growth. *Organizational dynamics*, 16(4): 5–21.
- Kaufmann, K. 2018. Navigating a new life: Syrian refugees and their smartphones in Vienna. *Information, Communication & Society*, 21(6): 882–898.
- Khatua, A. 2025. Policies for Refugee Mental Health Using Generative Agents. In *Companion Proceedings of the ACM on Web Conference 2025*, 1062–1066.
- Khatua, A.; and Nejdil, W. 2022. Endorsement Analysis of Migrant-related Deliberations on YouTube: Prior to and During 2022 Ukrainian crisis. In *Open Challenges in Online Social Networks*, 31–38. Association for Computing Machinery.
- Khatua, A.; and Nejdil, W. 2023. Why do we Hate Migrants? A Double Machine Learning-based Approach. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*, 1–10.
- Khatua, A.; Zagheni, E.; and Weber, I. 2023. Host-Centric Social Connectedness of Migrants in Europe on Facebook. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 1143–1147.
- Kim, J.; Sirbu, A.; Giannotti, F.; and Gabrielli, L. 2020. Digital footprints of international migration on Twitter. In *International symposium on intelligent data analysis*, 274–286. Springer.
- Kreis, R. 2017. #refugeesnotwelcome: Anti-refugee discourse on Twitter. *Discourse & communication*, 11(5): 498–514.
- Lillekroken, D.; Bye, A.; Halvorsrud, L.; Terragni, L.; and Debesay, J. 2024. Food for Soul—Older Immigrants' Food Habits and Meal Preferences After Immigration: A Systematic Literature Review. *Journal of Immigrant and Minority Health*, 1–31.
- Lingel, J.; Naaman, M.; and Boyd, D. M. 2014. City, self, network: transnational migrants and online identity work. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 1502–1510.
- Mayer, T.; and Zignago, S. 2011. Notes on CEPII's distances measures: The GeoDist database. Working Papers 2011-25, CEPII.
- Montreuil, A.; and Bourhis, R. Y. 2004. Acculturation orientations of competing host communities toward valued and devalued immigrants. *International Journal of Intercultural Relations*, 28(6): 507–532.
- Obradovich, N.; Özak, Ö.; Martín, I.; Ortuño-Ortín, I.; Awad, E.; Cebrián, M.; Cuevas, R.; Desmet, K.; Rahwan, I.; and Cuevas, Á. 2022. Expanding the measurement of culture with a sample of two billion humans. *Journal of the Royal Society Interface*, 19(190): 20220085.
- Rampazzo, F.; Bijak, J.; Vitali, A.; Weber, I.; and Zagheni, E. 2021. A framework for estimating migrant stocks using digital traces and survey data: An application in the United Kingdom. *Demography*, 58(6): 2193–2218.
- Ripani, G. 2023. Music in the lives of immigrant children: A review of literature. *Update: Applications of Research in Music Education*, 42(1): 51–60.
- Siapera, E.; Boudourides, M.; Lenis, S.; and Suiter, J. 2018. Refugees and network publics on Twitter: Networked framing, affect, and capture. *Social Media+ Society*, 4(1): 2056305118764437.
- Stewart, I.; Flores, R. D.; Riffe, T.; Weber, I.; and Zagheni, E. 2019. Rock, rap, or reggaeton?: assessing mexican immigrants' cultural assimilation using facebook data. In *The world wide web conference*, 3258–3264.
- Tadesse, B.; and White, R. 2010. Cultural distance as a determinant of bilateral trade flows: do immigrants counter the effect of cultural differences? *Applied Economics Letters*, 17(2): 147–152.
- Taras, V.; Kirkman, B. L.; and Steel, P. 2010. Examining the impact of Culture's consequences: a three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of applied psychology*, 95(3): 405.
- Van Oudenhoven, J. P.; and Ward, C. 2013. Fading majority cultures: The implications of transnationalism and demographic changes for immigrant acculturation. *Journal of Community & Applied Social Psychology*, 23(2): 81–97.
- Vieira, C.; Ribeiro, F.; Vaz de Melo, P. O.; Benevenuto, F.; and Zagheni, E. 2020. Using Facebook data to measure cultural distance between countries: the case of Brazilian cuisine. In *Proceedings of the web conference 2020*, 3091–3097.
- Vieira, C. C.; Lohmann, S.; Zagheni, E.; Vaz de Melo, P. O.; Benevenuto, F.; and Ribeiro, F. N. 2022. The interplay of migration and cultural similarity between countries: Evidence from Facebook data on food and drink interests. *Plos one*, 17(2): e0262947.
- Zagheni, E.; Garimella, V. R. K.; Weber, I.; and State, B. 2014. Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd international conference on world wide web*, 439–444.
- Zagheni, E.; Weber, I.; and Gummadi, K. 2017. Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 721–734.

Paper Checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **No**
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? **Yes**
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes**
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes**
 - (e) Did you describe the limitations of your work? **Yes**
 - (f) Did you discuss any potential negative societal impacts of your work? **NA**
 - (g) Did you discuss any potential misuse of your work? **NA**
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **NA**
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes**
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? **Yes**
 - (b) Have you provided justifications for all theoretical results? **NA**
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **NA**
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **NA**
 - (e) Did you address potential biases or limitations in your theoretical framework? **NA**
 - (f) Have you related your theoretical results to the existing literature in social science? **Yes**
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **NA**
3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? **NA**
 - (b) Did you include complete proofs of all theoretical results? **NA**
4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **NA**
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **NA**
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **NA**
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **NA**
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? **NA**
 - (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? **NA**
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
 - (a) If your work uses existing assets, did you cite the creators? **NA**
 - (b) Did you mention the license of the assets? **NA**
 - (c) Did you include any new assets in the supplemental material or as a URL? **NA**
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? **NA**
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **NA**
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR ? **NA**
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset ? **NA**
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
 - (a) Did you include the full text of instructions given to participants and screenshots? **NA**
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? **NA**
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **NA**
 - (d) Did you discuss how data is stored, shared, and de-identified? **NA**