Abstract

Extant literature has explored the social integration process of migrants settling in host communities. However, this literature typically takes a migrant-centric view, implicitly putting the burden of a successful integration on the migrant, and trying to identify the factors that lead to integration along various dimensions. In this paper, we flip this point of view by studying the attributes of natives that govern their propensity to form social ties with migrants. We do so by using anonymous and aggregate social network data provided by Facebook’s advertising platform. More specifically, we look at factors that influence the propensity for a likely-to-be non-Muslim Facebook user to have at least one social connection to a Facebook user who celebrates Ramadan. Given that, in the European context, following Islam is predominantly tied to a migration background, this gives us a lens into cross-cultural native-migrant connectivity. Our study considers demographic attributes of the host population, such as age, gender, and education level, as well as spatial variation across 30 European cities. Our findings suggest that young, educated, and male Facebook users are relatively more likely to build cross-cultural ties, compared to older, less educated, and female Facebook users. We also observe heterogeneity across the analyzed cities.

Social Connectedness of Migrants in Europe

Previous research pointed out that social connectedness, or the lack of social connectedness, depends on multiculturalism (Dijkstra, Geuïjen, and De Ruijter 2001). One of the drivers of multiculturalism in the European context is the inflow of migrants from Muslim-majority countries. Muslim migrants often face challenges regarding social acceptance in European societies, and debates around their social integration flame up regularly (Liebert, Siddiqui, and Goerzig 2020; De Coninck and Matthijs 2020). Prior studies noted that not only religiosity affects social contact but also social contact affects religiosity (Maliepaard and Schacht 2018). Additionally, the social connectedness of Muslim migrants into European society is becoming more challenging due to the rise of Islamophobia and anti-migrant sentiment (Ciftci 2012; Khatua and Nejdl 2022a,b). Muslims in Europe also face structural barriers to integration. For instance, Muslim migrants often come from poorer and politically volatile countries. As a result, many of them might have limited access to education, employment, and other opportunities, all factors that can play a role in their social integration (De Coninck and Matthijs 2020). Highly educated migrants may have more opportunities for employment and social advancement (Stempel and Alemi 2021; Adida, Lo, and Platas 2019), and may be less likely to face discrimination and prejudice. Those with lower levels of education may face greater challenges in terms of finding employment and integrating into their respective host societies. It has also been observed that Muslim women face greater challenges in terms of social integration and acceptance compared to Muslim men (Van Laar, Derks, and Ellemers 2013). This is often because Muslim women are more likely to wear traditional Islamic clothing, such as headscarves, which can make them more susceptible to discrimination.

Overall, various socioeconomic constraints make it difficult for Muslim migrants to fully participate in European society, and subsequently, they may become more isolated and marginalized. Our study probes host-centric social connectedness (HCSC) between hosts and migrants in the European context. Specifically, we assess which population groups in Europe are more socially connected to people who are likely Muslims.

Digital Studies of Social Connectedness

Given the interest of policymakers, academics, and the general public, in understanding and improving social integration, the literature has explored perceptions towards migrants (Khatua and Nejdl 2022b) and, subsequently, the socio-cultural integration of migrants. For example, prior studies explored economic integration through economic mobility or acceptance of migrants in the labor market (Chetty et al. 2022a,b) or cultural integration through food and musical interests on social media platforms like Twitter or Facebook (Kim et al. 2022; Stewart et al. 2019; Vieira et al. 2020). Using data from Twitter, researchers have also explored spatio-temporal language-based communication between migrants and hosts (Lamanna et al. 2018).

Methodologically, this study builds on prior works that used Facebook advertising data to study migrant assimilation in terms of interests expressed online (Dubois et al. 2018; Stewart et al. 2019). Migrants may struggle to settle
down in the host countries (Khatua and Nejdl 2022a). Social
connectedness on social media platforms with compatriots
from their home country, people originally from their new
country (locals), and also immigrants from other countries,
can help them to settle down (Herdadelen et al. 2016). A
stream of literature has probed social integration from the
perspective of migrants.

In contrast, this study flips this view and focuses on the
hosts. What are their characteristics that facilitate a cross-
cultural connectedness? In particular, which demographic
attributes are tied to a higher propensity of host-migrant
ties? Thus, our work complements the literature that stud-
ted demographic characteristics and political alignment of
host societies with respect to phenomena like Islamopho-
bia (Ogan et al. 2014). Our study also complements work
that looks at cross-lingual connections on Twitter (Kim et al.
2014) - as our work focuses on demographic attributes and
on cross-cultural connections. The work that is conceptually
most similar to ours is Bailey et al. (2022), as the authors
use Facebook data to study social integration for the spe-
cific case of Syrians in Germany, and to assess the effect of
integration courses on social integration. Our analysis has
a broader scope, considering Europe as a whole, and is not
limited to Syrians.

Both our study and Bailey et al. (2022) use online con-
nessedness to, ultimately, reason about real-world social ties.
While online ties might not necessarily accurately reflect off-
tline ties, recent work shows that ties, as observed on Face-
book, are causally linked to improved economic mobility
(Chetty et al. 2022a,b). In a similar vein, causal studies on
LinkedIn find a benefit of social network usage on job out-
comes (Wheeler et al. 2022). As such, even if our findings
only apply to online connectedness, they nevertheless speak
to real-world consequences.

**Data: Facebook’s Marketing API**

The key enabler of our study is an innovative use of so-
cial media advertising data from Facebook. This data comes
in the form of so-called audience estimates of how many
Facebook users match certain targeting criteria available to
advertisers. These anonymous and aggregate estimates are
available free of charge via an API, are accessible to ev-
everyone with a Facebook account, and have been used for a
number of socio-demographic studies before (Fatehkia et al.
2020a; Palotti et al. 2020; Zagheni, Weber, and Gummadi
2017).

The key differentiating factor in our work is the use of a
targeting attribute that relates to cross-cultural social con-
nectedness, namely: **Friends of people who have engaged
with Ramadan**. The advertising interface describes this tar-
geting option as “Friends of anyone that has high or medium
engagement with Ramadan month content. Excludes people
that engaged with Ramadan content.” In other words, us-
ing the advertising platform, we could obtain audience esti-
mates of how many likely-to-be non-Muslims have at least
one Facebook friend who is likely-to-be Muslim. We use
such data as evidence for the presence (or absence) of host-
migrant online ties, and we disaggregate our analysis along
different host attributes. The following host attributes chosen
for disaggregation are based on existing integration research
but are limited by what Facebook provides, and by what can
be obtained within the rate limits.

1. **Age**: 18 to 24; 25 to 34; 35 to 55; 56 and above [4
categories]
2. **Gender**: Male and Female [2 categories]
3. **Education**: Basic (such as High School, Some High
School, Undergrad, Associate Degree, etc.); Advanced (such
as Professional Degree, Master’s Degree, Doctorate Degree,
etc.); Unknown/unspecified [3 categories]
4. **Operating system**: iOS users vs. non-iOS users [2 cat-
egories]
5. **Local language**: Not for disaggregation, but to identify
the hosts: we limited the audience selection to speakers of
the respective local language (German for cities in Germany,
Italian for cities in Italy, and so on). ³
6. **Friends with upcoming birthdays**: The likelihood to
have a Muslim friend on Facebook also depends on the size
of the person’s online social network. While we cannot ob-
sERVE the absolute size directly, we can observe whether a
user has a friend on Facebook with a birthday coming up
in the next week. The probability of this happening is tied
to the network size and can be used to indirectly observe
variation thereof (Gil-Clavel, Zagheni, and Bordone 2022).
Thus, we use it for normalizing our dependent variable and
controlling the size of the host’s online social network.
7. **Exposure to AR/TR**: The propensity to be friends with
Muslims also depends on whether hosts have exposure to
Muslim migrants in their city: someone in a city with only
a small Muslim population would naturally be less likely
to befriend a Muslim. Unfortunately, we could not obtain
city-level Muslim population estimates for all 30 cities in
our study. Thus, as a proxy for this (missing) exposure vari-
able, we obtained city-level Facebook audience estimates for
the Arabic (AR) and Turkish (TR)-speaking users as a per-
centage of total Facebook users for that specific city. Impor-
tantly, this Facebook-derived variable is the only one that is
not for the individual hosts but for their city.

For our study, we collected data for 30 cities from seven
European Countries (see Figure 1 for details), picking the
most populous cities in each country, ranging from six cities
for Germany to three for the Netherlands. This selection
still included some comparably small cities such as Bil-
bao (Spain) or Goteborg (Sweden). For the smaller cities
and certain combinations of attributes, we had sparsity chal-
lenges as Facebook’s advertising API does not return audi-
ence estimates of monthly active users (MAU) smaller than
1,000. For our analysis, we decided to keep only the non-
sparse (> 1,000) estimates, meaning that the number of de-

³Facebook’s advertising platform currently does not support the
ad targeting of non-binary users.

²Previous work has found that, across numerous geographic con-
texts, using an iOS, i.e. Apple device to access Facebook is
generally a proxy for higher wealth (Fatehkia et al. 2020b, 2022).
³While relying on language is simplistic, the targeting attribute
used explicitly “[e]xcludes people that engaged with Ramadan con-
tent”. We believe that the combination of these two provides a good
enough proxy for the native host population.
Relative to the baseline (age 18-24, female, basic education, non-iOS) with positive coefficients indicating higher connectedness.

Table 1: Effects of Demographic and Spatial Characteristics on Host-Centric Social Connectedness (HCSC). Coefficients are relative to the baseline (age 18-24, female, basic education, non-iOS) with positive coefficients indicating higher connectedness.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.75***</td>
<td>0.74***</td>
<td>0.73***</td>
<td>0.74***</td>
<td>0.68***</td>
<td>0.80***</td>
<td>0.70***</td>
<td>0.87***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Age: 25 to 34</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01*</td>
<td>0.01</td>
<td>0.02*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Age: 35 to 55</td>
<td>−0.08***</td>
<td>−0.07***</td>
<td>−0.08***</td>
<td>−0.08***</td>
<td>−0.09***</td>
<td>−0.07***</td>
<td>−0.07***</td>
<td>−0.07***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Age: 56 and above</td>
<td>−0.20***</td>
<td>−0.20***</td>
<td>−0.20***</td>
<td>−0.21***</td>
<td>−0.20***</td>
<td>−0.19***</td>
<td>−0.20***</td>
<td>−0.20***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Gender: Male</td>
<td>0.02***</td>
<td>0.02***</td>
<td>0.02***</td>
<td>0.02***</td>
<td>0.02***</td>
<td>0.02***</td>
<td>0.02***</td>
<td>0.02***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Education: Advanced</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Education: Unknown</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>OS: iOS users</td>
<td>−0.01</td>
<td>−0.01</td>
<td>−0.02***</td>
<td>−0.02***</td>
<td>−0.01***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Exposure to AR/TR</td>
<td>0.01***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City-dummies</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>R²</td>
<td>0.35</td>
<td>0.36</td>
<td>0.37</td>
<td>0.37</td>
<td>0.46</td>
<td>0.68</td>
<td>0.69</td>
<td>0.78</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.35</td>
<td>0.36</td>
<td>0.37</td>
<td>0.37</td>
<td>0.46</td>
<td>0.67</td>
<td>0.69</td>
<td>0.77</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>1075</td>
<td>1075</td>
<td>1075</td>
<td>1075</td>
<td>1075</td>
<td>1075</td>
<td>1075</td>
<td>1075</td>
</tr>
</tbody>
</table>

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

Table 1: Effects of Demographic and Spatial Characteristics on Host-Centric Social Connectedness (HCSC). Coefficients are relative to the baseline (age 18-24, female, basic education, non-iOS) with positive coefficients indicating higher connectedness.

Figure 1: Country-wise HCSC Scores

mographic combinations used per city varies and is typically smaller for cities with a smaller population.

The monthly active user audience estimates were collected in two rounds, January and March of 2022, from Facebook’s advertising API. We averaged the audience estimates from the two rounds of data collection to improve robustness and limit temporal fluctuation. For a demographic sub-group of interest, such as “18-24 years old German-speaking women with higher education, using iOS devices, and living in Berlin”, this variable captures the propensity to have at least one likely-to-be Muslim friend on Facebook. This propensity is normalized by the propensity to have at least one friend with an upcoming birthday in the next week. Thus, HCSC, our dependent variable, is # of people with a Ramadan-celebrating friend / # of people with a friend with an upcoming birthday in the next week.

Figure 1 reports the country-level HCSC score, i.e, the average of city-level scores for that specific country. We find variation across the seven considered countries. For instance, the score of the Netherlands is 0.80 (highest), whereas it is 0.56 for Spain (lowest). In the following section, we probe deeper using linear regression to understand the relationship between HCSC Score and demographic attributes, as well as city-level variation.

Findings

Age Effects: In all of our models (see Table 1), age-related effects of the 35-55 and 56+ years groups, relative to the 18-24 baseline group, are negative and statistically significant (p < 0.001). Furthermore, these effect sizes (≈ −0.08 and ≈ −0.20 for 35-55 and 56+, respectively) are the biggest ones for demographic variables, comparable to some between-city variation (see Figure 2). Variation in age alone also ex-
Spatial Effects: Finally, M6 to M8 control the spatial effects through dummy variables. Interestingly, the Adj. $R^2$ went up significantly from 0.46 (in M5) to 0.67 (in M6) after controlling for country effects. However, the improvement is marginal between M6 (with country dummies) and M7 (with exposure effects + country dummies).

Finally, we have incorporated the city-level dummy variables in M8 and reported the city-level effects graphically in Figure 2, showing strong variation across the analyzed cities. Correspondingly, the Adj. $R^2$ went up significantly from 0.69 (in M7) to 0.77 (in M8), indicating existing city-level effects beyond exposure. Note that we cannot simultaneously include both City-Level Exposure Effects and City Dummies in our models due to multicollinearity. Interestingly, only city dummies report an Adj. $R^2$ of 0.41.

Ethical Considerations
Anonymous and aggregate data were obtained through Facebook’s Marketing API. Given the minimum group size of 1000, any individual re-identification risk is minimal. However, there is a risk of group-level harm by mapping vulnerable populations, such as those of a particular faith. To mitigate this risk, Facebook removed targeting attributes related to religion and other sensitive attributes, including the one used in this study 5. Note, however, that our study does not target Muslim migrants themselves but ‘natives’ non-Muslim in the respective countries, limiting the potential group harm. Still, the removal of the targeting attribute of “friends of people who have engaged with Ramadan” limits the reproducibility. Given the sensitivity of the topic, we commit to sharing our data with other researchers upon request.

Limitations and Conclusion
Our study addresses the issue of integration in an innovative way. However, threats to the validity of this type of analysis include (i) ecological fallacy, due to using group-level data, and (ii) invalid constructs, due to reliance on proxy measures. As a plausibility check for our results, we note that certain observed results are in line with existing findings, such as a higher level of social integration in the Netherlands than in Italy or Spain (Bell, Valenta, and Strabac 2021). Furthermore, we also collected data for Turkey, a Muslim-majority country, where we found a country-level HCSC score of almost 1, higher than the highest value for all European countries. This indicates that our target measure satisfies at least basic plausibility checks.

We are aware that celebrating Ramadan relates to religious or cultural behavior and not necessarily a migration background. However, in the context of Western Europe, most Muslims have a migration background (Bailey et al. 2022), leading us to emphasize the host-migrant connection angle, in addition to the cross-cultural angle.

In conclusion, by studying online connectedness and taking a host-centric approach, this paper contributes to a more nuanced debate around integration processes in Europe.

---

Acknowledgments

This research was partially funded by the Federal Ministry of Education and Research (BMBF), Germany under the project LeibnizKILabor with grant No. 01DD20003.

References


