3

SOCIAL MEDIA

DISPLAYING SOCIAL CONNECTIONS

MEASURING MIGRANT STOCKS AND FLOWS

SAMPLING MIGRANTS AND DIASPORAS

# AUGMENTING MIGRATION STATISTICS USING SOCIAL MEDIA

*Jisu Kim,[1] Emilio Zagheni[2] and Ingmar Weber[3]*

## Introduction

Migration is a key driver of demographic and social change, as well as an often hotly debated political issue. Over the past years, the phenomenon has attracted attention from policymakers, researchers, as well as the general public, particularly due to increasing numbers of forcibly displaced individuals. At the end of 2019, about 272 million people migrated (DESA, 2019a), and about 79.5 million people were forcibly displaced from their homes, either internally or internationally, due to conflicts, generalized violence or human rights violations.[4] These numbers have continued to increase (DESA, 2019b), up until the outbreak of COVID-19 which seems to have halted this growth, although there is speculation that we may observe a spike in the number of migrants once the travel restrictions are eased (O'Brien and Eger, 2020). In order to cope with the complexity of the matter, the need for up-to-date and rich data to better monitor and manage the situation has become clear. However, traditional migration data such as census, register and survey data that researchers mainly rely on have a number of limitations and imperfections (Sîrbu et al., 2020).

First, the definitions used may be inconsistent across different countries. For example, the reported number of people immigrating to Spain from Italy in 2015 was 17,350 according to official data from the Spanish National Statistics Institute. The same number was estimated to be only 5,003 according to official data from the Italian National Institute of Statistics.[5] This discrepancy is in part related to the varying definitions of "migrant" and the different ways of collecting data. It is also due to the fact that there are limited incentives for emigrants to declare their departure in their origin country, and more incentives for immigrants to register in the country of destination. Apart from inconsistencies, traditional data are often time-consuming, costly to collect, and published with delay. In extreme circumstances, there are no data at all.

[1] Jisu Kim is a research scientist at Max Planck Institute for Demographic Research. Her research focuses on the intersection of migration sciences, the economics of migration, complex social networks, statistical models and data-driven algorithms. Jisu holds a PhD in Data Science from the Scuola Normale Superiore in Italy.

[2] Emilio Zagheni is Director of the Max Planck Institute for Demographic Research in Rostock, Germany, and Head of the Laboratory of Digital and Computational Demography. He is best known for his pioneering work on using web and social media data for studying migration processes and for his role in developing the field of digital and computational demography.

[3] Ingmar Weber is an Alexander von Humboldt Professor for Artificial Intelligence at Saarland University. Previously he was Research Director for Social Computing at the Qatar Computing Research Institute. He studied and worked at the University of Cambridge, the Max Planck Institute for Informatics, the Ecole Polytechnique Fédérale de Lausanne and Yahoo Research.

[4] More information is available at www.unhcr.org/globaltrends2019/.

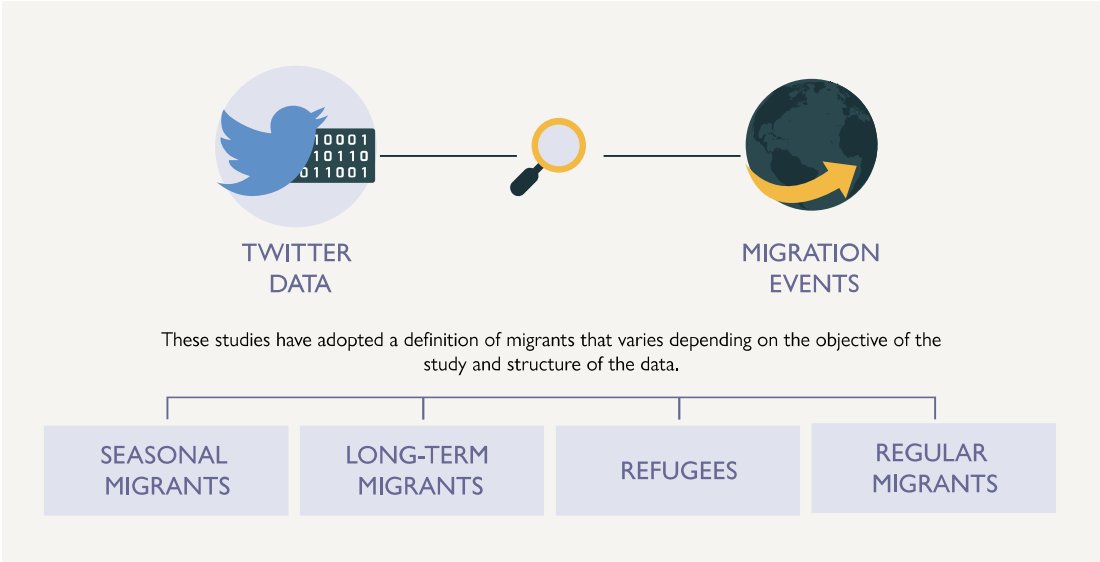
[5] More information is available at https://ec.europa.eu/eurostat/home.

As we live in the digital age, a vast amount of passively collected data exists from new sources, such as call detail records, social media platforms and other Internet services. Some of them are freely available and enable us to gather real-time data. Furthermore, these sources provide large-scale data on a granular level, which enables scientists to study phenomena that are hard to analyse using more coarse-grained data. Aside from potential advantages such as the low cost, temporal and spatial granularity, and real-time availability, these non-traditional data sources come with a range of challenges, particularly in relation to selection biases and ethical issues.

This chapter focuses on the potential value of data from two social media platforms: (a) geotagged tweets and (b) Facebook ads audience estimates – to demonstrate how these data can be used for migration statistics. The declared goal here is to enhance, not replace, traditional migration statistics. The chapter is divided into two main sections; the first section focuses on Twitter and geotagged tweets, whereas the second section concentrates on Facebook ads audience estimates. Each section discusses how the data can be accessed, provides details on the drawbacks of the data, and describes how such data have been used to produce estimates and enhance migration statistics. We conclude the chapter with a summary of overall pros and cons of social media data and the potential of these data for exploring additional dimensions of migration processes that are difficult to quantify with traditional sources only.

## Accessing and using Twitter data to measure migration

In the past years, several studies have used Twitter data to study migration events (Fiorio et al., 2017; Hausmann et al., 2018; Hawelka et al., 2014; Huang et al., 2020; Lamanna et al., 2018; Kim et al., 2020; Mazzoli et al., 2020; Zagheni et al., 2014). These studies have adopted a definition of "migrant" that varies depending on the objective of the study and structure of the data, and have focused on various groups of migrants, such as seasonal migrants, long-term migrants, refugees or regular migrants. Studies commonly try to pinpoint a country of residence by looking at the most frequent location of tweets for a period considered. For instance, Zagheni et al. (2014), Hawelka et al. (2014), and Kim et al. (2020) all define "country of residence" as the country where a user tweets for the most part over a period of four months (Zagheni et al., 2014) or a year (Kim et al., 2020). According to Zagheni et al. (2014), outmigration events are detected when there is a change in the country of residence.



On the other hand, migration events are detected, according to Kim et al. (2020), when the nationality identified is different from the country of residence – where "nationality" is defined by "linguistic and social connections to a migrant's country of origin". As reported by Hawelka et al. (2014), a mobility pattern is detected once a country other than the country of residence appears in tweets. However, different from the other works, Hawelka et al.'s article (2014) studied global mobility and was able to detect seasonal mobility patterns which fall under the definition of "visitors" rather than migrants, under the definition set by DESA's *Recommendations on Statistics of International Migration* (1998). Similarly, Mazzoli et al. (2020) have observed routes of migrants and refugees during the migration crisis in the Bolivarian Republic of Venezuela by detecting a change in location during the time window of observation. The emigration identification strategy is somewhat similar also in Hausmann et al.'s working paper (2018). Although they were not focusing on migrants, strictly speaking, Huang et al. (2020) showed that Twitter data can be a good data source to study how mobility patterns have changed after international efforts to reduce mobility were implemented during the middle of the COVID-19 pandemic.

A challenge common to all social media-based studies, in particular those using Twitter data, is how to validate the results. In other words, how can one make sure that the results are not biased and that, ideally, the findings can be generalized to the wider population, not only a small subset of Twitter users? As the studies mentioned above have shown, there are numerous ways to deal with biases. For instance, Zagheni et al. (2014) employed the difference-in-differences technique to control for different Twitter usage across countries by assuming that, in the short term, differences in bias across countries may be constant. Other studies have validated their results by comparing the predicted results with official statistics (Hawelka et al., 2014; Kim et al., 2020; Mazzoli et al., 2020). In a recent contribution, Hsiao et al. (2020) showed that once spatial and temporal bias structures are statistically modelled from Twitter data, other data sources such as survey data can be combined to develop a joint model. In addition, the authors concluded that not only did the joint model correct the upward bias in Twitter data, but it also outperformed the accuracy level of survey data alone.

In terms of data collection, Twitter is a freely available data source that can be accessed using an application programming interface (API).[6] The two main methods of accessing data using the Twitter API are through the search API and the streaming API. Both APIs return data in JSON format, called "objects",[7] which are easy to store and manipulate.

The search API enables the collection of publicly available tweets and profiles of users. The search can be done on a specific user through either the user ID or the username. This returns a user object that contains information about an individual's profile, such as when the account was created; the number of tweets, followers and friends; as well as the location that the user has declared to be in. Otherwise, the search can be run on specific keywords or geolocations embedded in tweets. The geolocations can be specified – a country, place, bounding box, or within a radius of 0.01 km up to about 40 km from specific coordinates.[8]

[6] More information is available at https://developer.twitter.com/en/docs/twitter-api.

[7] More information is available at www.json.org/json-en.html.

[8] More information is available at https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location.

On the other hand, the streaming API allows us to gather random samples from 1 per cent of all new public tweets in near real time. The streaming API also allows us to specify filter criteria (e.g. keywords, geolocations, and user IDs or usernames). However, unlike the search API, it returns tweets matching the filter criteria as soon as matching tweets are created. The returned content of matching tweet objects includes the tweet text, location information (where present), the language in which the tweet was written when it was created, and additional information, such as whether the tweet was part of a thread. It also contains the entity object which lists tweet details such as hashtags, URLs and mentioned IDs. To collect small amounts of data, instead of using the APIs, there are websites to search for tweets (e.g. on Twitter directly),[9] or where detailed searches for particular users can be issued (e.g. on followerwonk).[10]

**Figure 1. An example of a search on followerwonk.com for "IOM GMDAC"**


After having obtained the data, there are few notes of caution to consider. First, only a small percentage of tweets come with geolocations based on the user opting in to share their exact position. For instance, Morstatter and Liu (2017) showed that only about 3.2 per cent of tweets from the streaming API are geotagged. This means that any given user is unlikely to have geotagged tweets and, correspondingly, challenges related to self-selection bias need to be addressed.

Second, it requires effort to clean and process the data. Often, the tweets are not directly usable as they are "noisy" and/or incomplete. For instance, tweets contain repeated characters (e.g. "wooooooow"), typos, or Internet slangs that are not familiar to everyone, and these pose challenges to standard natural language processing (NLP) tools. Some tweets may also be incomplete in that they require additional context, such as a thread of tweets, to make sense of them. Cleaning and removing data may result in a considerable loss of information. Further, bots or spam accounts introduce additional data quality issues. It is also important to make sure that identifying migration events is not confounded with misleading activities.

Another limitation is that Twitter does not provide user attributes such as education or income level, which are often helpful for more in-depth migration studies. Nevertheless, certain characteristics, such as age, ethnicity or sex, can often be inferred with reasonable accuracy using the profile image (Zagheni et al., 2014; Huang et al., 2014).

[9] More information is available at https://twitter.com/explore.

[10] More information is available at https://followerwonk.com.

Lastly, there are privacy issues. It is vital to make sure that no personal information obtained from the data is published even if Twitter data are openly available. A proper infrastructure is needed where data can be safely stored in a secured server. Additionally, to comply with Twitter's Terms of Service, only "dehydrated" data sets can be shared for research or archival purposes, except for small collections not exceeding 50,000 tweets.[11] This requires data to be in the form of unique IDs which can then be "rehydrated" – in other words, restored to the original data.[12] This gives the user a chance to opt out of subsequent studies by deleting their tweet/account. It should also be noted that while most Twitter content is public and accessible to anybody, an individual user might not expect researchers to algorithmically collect and analyse their tweets. How to best address these expectations of data use, which are separate from legal considerations, remains a challenge, with answers depending on the specific context.

## Accessing and using Facebook ads audience estimates to measure migration

Given the size of Facebook, with more than 2.7 billion users in September 2020,[13] researchers have also tried to tap into data from the social networking site, in particular through its advertising platform. Here existing works have shown that Facebook can indeed be a good alternative data source to study migration phenomena (Alexander et al., 2019; Gendronneau et al., 2019; Palotti et al., 2020; Spyratos et al., 2018, 2019; Zagheni et al., 2017). The studies have all shown that stocks or flows of migrants identified through Facebook data are indeed well correlated with official statistics in various cases. For instance, Zagheni et al. (2017) found that 94 per cent of the variability in data from the American Community Survey (ACS) is explained by Facebook's estimates. In other words, the relative trends observed in Facebook's estimates with, say, a lot of migrants from country X in location A, but relatively few in location B, are very close to official statistics. Palotti et al. (2020) also found a correlation coefficient of 0.99 when comparing estimates for the number of Venezuelan migrants and refugees across 17 host countries as reported in a 2019 R4V[14] report with Facebook estimates for the same group of countries.

Apart from the good spatial similarity to ground truth data, the real-time value of Facebook data has also been demonstrated in the setting of an exodus after a natural disaster or economic crisis (Spyratos et al., 2019; Palotti et al., 2020; Alexander et al., 2019). Spyratos et al. (2019) and Palotti et al. (2020) looked at the case study of Venezuelan migrants and refugees after the economic crisis to study their distribution in neighbouring countries. Both groups were able to successfully detect an increase in Venezuelan migrants and refugees in surrounding countries after the crisis hit. On the other hand, Alexander et al. (2019) analysed outmigration from Puerto Rico in the aftermath of Hurricane Maria in 2017. As these studies have shown, Facebook data allow researchers to obtain more recent data, enabling us to carry out research quickly, especially during humanitarian crises. Another added value of Facebook is that collecting on-the-ground data is possible even during a period when in-person data collection is difficult to pursue (Perrotta et al., 2021; Pötzschke and Braun, 2017), not to mention that the cost and logistical effort are considerably less compared to traditional approaches.

[11] More information is available at https://developer.twitter.com/en/developer-terms/agreement-and-policy.

[12] More information is available at https://scholarslab.github.io/learn-twarc/06-twarc-command-basics#dehydrated-and-rehydrated-data-sets.

[13] More information is available at https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Third-Quarter-2020-Results/default.aspx (accessed September 2020).

[14] More information is available at https://r4v.info/.

While only a limited number of studies have been conducted on the bias in Twitter data, more evidence exists on the bias in Facebook data since Facebook-derived estimates can be compared with traditional statistics directly. Since information on individuals' socioeconomic status is also available in most countries, some studies have taken into account the differences in penetration rates by age and sex across countries to fully understand the origin of the bias.

For instance, Spyratos et al. (2019) considered differences in penetration rates by age and gender both in the country of current residence and in the country of previous residence. Palotti et al. (2020) found that Facebook tends to overestimate the number of migrants and refugees compared to official statistics – which might themselves be underestimated. The same study also showed evidence that Facebook data might underestimate the number of migrants and refugees in less wealthy areas. The measurement of wealth was inferred by looking at the prices of devices used to access Facebook. Zagheni et al. (2017) observed that Facebook generally overestimates migrant stocks in younger age groups and underestimates stocks in older age groups. Last but not least, Ribeiro et al. (2020) analysed biases across several dimensions including race, income level and age – at different geolocation levels including country, state and city. They found that bias exists mainly towards young people and women but also towards college graduates. However, Facebook tends to capture trends relatively similar to survey data when looking at individuals with high school or graduate school levels of education and different income levels.

The Facebook advertising platform[15] was originally designed for purposes of targeted marketing to allow advertisers to select and reach a specific audience. However, this tool has also shown its utility in studying migration. Different from Twitter, Facebook provides explicit criteria that enable advertisers to selectively show ads to users that are likely migrants. Concretely, Facebook provides options to selectively show ads to any Facebook user who "used to live in [country X]" (expats), "lives abroad" and "recently moved". Most studies, described further below, use the targeting criterion "used to live in [country X]" as a proxy for migration history. The Facebook advertising platform then provides an estimate of the size of the audience matching the chosen criteria. These estimates include numbers for both monthly and daily active users. To protect user privacy, the returned estimates are rounded to two significant digits. Furthermore, user groups smaller than 1,000 users are indistinguishable from 0, as FB never returns estimates of monthly active users smaller than 1,000.

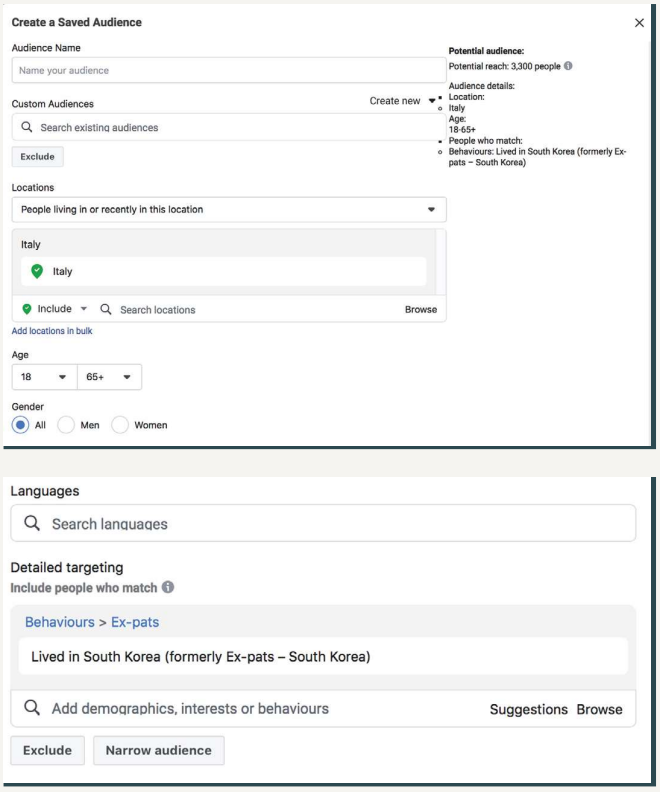
One of the key challenges for researchers is that Facebook does not disclose the methodology used to produce these estimates. For example, "expats" are briefly defined as users "who used to live in [country X] who now live abroad", but how previous countries of residence are inferred is unclear. Furthermore, this vague definition leaves us to wonder whether it closely corresponds to the definition of "migrant", as set in the United Nations 1998 Recommendations or in the revised version by the United Nations Statistical Commission. The first states that an international migrant is "any person who changes his or her country of usual residence" (DESA, 1998:9). On the other hand, the revised definition specifies that "a person who has changed his or her country of residence and established new residence in the country within a given year" (UNSD, 2020:13) is considered as an international migrant.

[15] More information is available at www.facebook.com/business/ads.

However, from the work of Herdağdelen et al. (2016), done by researchers at Facebook, we are able to get clues about the features that likely contribute to the identification of expats on the site. They first identify individuals who are now living in the United States of America but who "specify a hometown outside of the United States", their country of interest. To further reduce noise due to users not stating their true hometowns, the researchers have also considered social network information to determine if a particular hometown is plausible. At the time of writing, Facebook supports 89 countries of origin of expats. Aside from migration estimates, the Facebook ads platform provides users' demographic information such as their self-declared level of education, field of study, schools/universities attended, field or industry, and also behaviours and interests. Behaviours include various categories such as "frequent travellers" and likely engagements of individuals in conservative, liberal or moderate political contents in the United States.

A particularly useful targeting attribute is the device type and the operating system used to access Facebook. Previous works have shown that, as a first approximation, Apple iOS devices are more common among population groups with higher disposable incomes (Fatehkia et al., 2020; Palotti et al., 2020). The targetable interests include various domains, such as foods/drinks, fashion, entertainment, sports and technology – and others which are observed through pages that users "liked". Not all targeting attributes are available in all countries, with the richest set of targeting capabilities found in the United States. All of these data can be accessed using either the Facebook Ads Manager or Marketing API.[16]

**Figure 2. An example of a Facebook ads platform audience search**




The Facebook advertising platform was originally designed for purposes of targeted marketing to allow advertisers to select and reach a specific audience. However, this tool has also shown its utility in studying migration.

[16] More information is available at https://developers.facebook.com/docs/marketing-apis/.

As with Twitter, using Facebook as a data source also comes with several drawbacks. First of all, there are bots and fake accounts that may interfere with how Facebook identifies expats. According to Rosen (2020), there were about 1.5 billion fake accounts in the second quarter of 2020. Fortunately, most of these accounts are detected and removed almost immediately after their creation so that they account for approximately "only" 5 per cent of the worldwide monthly active users (MAU) at any given time (ibid.). Under Facebook's policy, misrepresentation of one's identity on a profile classifies as having a fake account.[17] This includes any personal profiles created to represent pets, businesses or organizations – or profiles managed by multiple users. Owning multiple accounts is also not allowed by Facebook's policy. As for spam, there were about 1.4 billion spam activities detected on Facebook during the same period. Spam activities involve inflating posts, likes or shares that may mislead audiences typically for financial gains.[18]

While with Twitter, researchers have the flexibility to implement their own algorithm to detect bots and fake accounts, this is not possible with Facebook as no anonymized individual data is shared and the exact algorithms employed to estimate the size and composition of user groups are not fully disclosed. In addition, we can extract audience estimates of Facebook only at the point of querying the API. While this makes perfect sense for advertisers, who are not interested in how many users they could have reached, say, one year ago, this limits research on historical migration trends.

## Conclusion

There are both advantages and drawbacks in using social media data to study migration events. Twitter and Facebook data are freely available through public APIs, but Twitter requires considerably more API calls as individual – not aggregate – data needs to be collected. Collecting large amounts of Twitter data is also becoming more and more challenging due to increased scrutiny, resulting in regular changes in rate limits. However, once obtained, Twitter data are relatively easier to interpret. Facebook data are more difficult to interpret as the exact algorithms used to extract the estimates are not disclosed. Also, the methods periodically change, resulting in discontinuities that need to be accounted for. With Twitter data, longitudinal studies are possible as Twitter supports the collection of historical data. This is not feasible with real-time Facebook advertising data.

Although the privacy issues were not treated in depth in this chapter, they constitute one of the crucial limitations of Twitter data. Twitter is an open data platform, but users are not necessarily aware of researchers analysing their conversations and activities. To manage privacy issues, it is necessary to take security measures such as pseudonymization, or anonymization, to prevent reidentification of individuals and protect personal data. Furthermore, it is essential for researchers to secure the collected data to ensure that both raw and processed data are used responsibly and ethically. This would involve storing data in a secured server or limiting access to them. All of these privacy issues are of a lesser concern with Facebook data as the site provides only estimates of migrants at an aggregate level. At the same time, questions of potential group harm are more pronounced as it is relatively easy to identify, or rather enumerate, potentially vulnerable groups, based on variables such as income or ethnicity.

[17] More information is available at www.facebook.com/communitystandards/misrepresentation.

[18] More information is available at www.facebook.com/communitystandards/spam.

The quality of migration data varies vastly from one country to another, and for those with low-quality data, the need for better data has been pointed out. In such data-poor environments, social media data could help augment traditional ways of data collection to create better migration statistics. However, for countries with strong and real-time registration systems, social media is unlikely to add meaningful value in terms of providing better statistics. But even in these settings, innovative data might be able to complement existing data by adding new dimensions. For example, it seems promising to augment the Integration Index that is traditionally developed using civil registration data (Bansak et al., 2018). This index could potentially be improved by adding new insights from social media data, such as spatial and cultural aspects of integration (Dubois et al., 2018; Mazzoli et al., 2020; Stewart et al., 2019) – where, in addition to traditional measures of culture (such as languages, religion or marital status), "likes" or interests of individual users, as they relate to different aspects of life, can be studied through Facebook data.

# REFERENCES*

Alexander, M., K. Polimis and E. 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.

Bansak, K., J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence and J. Weinstein

2018 Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329.

Dubois, A., E. Zagheni, K. Garimella and I. Weber

2018 Studying migrant assimilation through Facebook interests. Accepted as a short paper at the International Conference on Social Informatics. Springer, pp. 51–60.

Fatehkia, M., I. Tingzon, A. Orden, S. Sy, V. Sekara, M. Garcia-Herranz and I. Weber

2020 Mapping socioeconomic indicators using social media advertising data. *EPJ Data Science*, 9:22.

Fiorio, L., G. Abel, J. Cai, E. Zagheni, I. Weber and G. Vinué

2017 Using Twitter data to estimate the relationship between short-term mobility and long-term migration. In: *WebSci '17: Proceedings of the 2017 ACM on Web Science Conference*. Association for Computing Machinery (ACM), New York, pp. 103–110.

Gendronneau, C., A. Wiśniowski, D. Yildiz, E. Zagheni, L. Fiorio, Y. Hsiao, M. Stepanek, I. Weber, G. Abel and S. Hoorens

2019 *Measuring Labour Mobility and Migration Using Big Data: Exploring the Potential of Social-Media Data for Measuring EU Mobility Flows and Stocks of EU Movers*. European Commission, Brussels. Available at www.rand.org/pubs/external_publications/EP68038.html.

Hausmann, R., J. Hinz and M.A. Yildirim

2018 Measuring Venezuelan emigration with Twitter. Kiel Working Paper No. 2106. Kiel Institute for the World Economy.

Hawelka, B., I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos and C. Ratti

2014 Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271.

Herdağdelen, A., B. State, L. Adamic and W. Mason

2016 The social ties of immigrant communities in the United States. In: *Proceedings of the Eighth ACM Conference on Web Science*. ACM, New York, pp. 78–84.

** All hyperlinks were working at the time of writing this report.*

Hsiao, Y., L. Fiorio, J. Wakefield and E. Zagheni
2020 Modeling the bias of digital data: An approach to combining digital and survey data to estimate and predict migration trends. MPIDR Working Paper WP-2020-019. Max Planck Institute for Demographic Research, Rostock, Germany.

Huang, X., Z. Li, Y. Jiang, X. Li and D. Porter
2020 Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLOS ONE*, 15(11):e0241957.

Huang, W., I. Weber and S. Vieweg
2014 Inferring nationalities of Twitter users and studying inter-national linking. In: *HT '14: Proceedings of the 25th ACM Conference on Hypertext and Social Media*. ACM, New York, pp. 237–242.

Kim, J., A. Sîrbu, F. Giannotti and L. Gabrielli
2020 Digital footprints of international migration on Twitter. In: *International Symposium on Intelligent Data Analysis*. Springer, pp. 274–286.

Lamanna, F., M. Lenormand, M.H. Salas-Olmedo, G. Romanillos, B. Gonçalves and J.J. Ramasco
2018 Immigrant community integration in world cities. *PLOS ONE*, 13(3):e0191612.

Mazzoli, M., B. Diechtiareff, A. Tugores, W. Wives, N. Adler, P. Colet and J.J. Ramasco
2020 Migrant mobility flows characterized with digital data. *PLOS ONE*, 15(3):e0230264.

Morstatter, F. and H. Liu
2017 Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media,* 1:1–13.

O'Brien, M.L. and M.A. Eger
2020 Suppression, spikes, and stigma: How COVID-19 will shape international migration and hostilities toward it. *International Migration Review*, 55(3):640–659.

Palotti, J., N. Adler, A. Morales-Guzman, J. Villaveces, V. Sekara, M. Garcia Herranz, M. Al-Asad and I. Weber
2020 Monitoring of the Venezuelan exodus through Facebook's advertising platform. *PLOS ONE*, 15(3):e0230455.

Perrotta, D., A. Grow, F. Rampazzo, J. Cimentada, E. Del Fava, S. Gil-Clavel and E. Zagheni
2021 Behaviors and attitudes in response to the COVID-19 pandemic: Insights from a cross-national Facebook survey. *EPJ Data Science*, 10:17.

Pötzschke, S. and M. Braun
2017 Migrant sampling using Facebook advertisements: A case study of Polish migrants in four European countries. *Social Science Computer Review*, 35(5):633–653.

Ribeiro, F.N., F. Benevenuto and E. Zagheni

2020 How biased is the population of Facebook users? Comparing the demographics of Facebook users with census data to generate correction factors. In: *WebSci '20: 12th ACM Conference on Web Science*. ACM, New York, pp. 325–334.

Rosen, G.

2020 Community standards enforcement report: August 2020. Facebook. Available at https://about.fb.com/news/2020/08/community-standards-enforcement-report-aug-2020/.

Sîrbu, A., G. Andrienko, N. Andrienko, C. Boldrini, M. Conti, F. Giannotti, R. Guidotti, S. Bertoli, J. Kim, C.I. Muntean, L. Pappalardo, A. Passarella, D. Pedreschi, L. Pollacci, F. Pratesi and R. Sharma

2020 Human migration: The big data perspective. *International Journal of Data Science and Analytics*, 11:341–360.

Spyratos, S., M. Vespe, F. Natale, I. Weber, E. Zagheni and M. Rango

2018 *Migration Data Using Social Media: A European Perspective*. Publications Office of the European Union.

2019 Quantifying international human mobility patterns using Facebook network data. *PLOS ONE*, 14(10):e0224134.

Stewart, I., R.D. Flores, T. Riffe, I. Weber and E. Zagheni

2019 Rock, rap, or reggaeton?: Assessing Mexican immigrants' cultural assimilation using Facebook data. Proceedings of the World Wide Web Conference, pp. 3258–3264.

United Nations Department of Economic and Social Affairs (DESA)

1998 *Recommendations on Statistics of International Migration*. Revision 1. New York. Available at https://unstats.un.org/unsd/publication/seriesm/seriesm_58rev1e.pdf.

2019a *International Migration 2019: Report*. ST/ESA/SER.A/438. Population Division, New York. Available at www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/InternationalMigration2019_Report.pdf.

2019b *International Migration 2019: Wall Chart*. ST/ESA/SER/A/431. Population Division, New York. Available at www.un.org/en/development/desa/population/migration/publications/wallchart/docs/MigrationStock2019_Wallchart.pdf.

United Nations Statistical Commission (UNSD)

2020 Report of the Secretary-General on migration statistics. E/CN.3/2021/11. Available at https://unstats.un.org/unsd/statcom/52nd-session/documents/2021-11-MigrationStats-EE.pdf.

Zagheni, E., V.R.K. Garimella, I. Weber and B. State

2014 Inferring international and internal migration patterns from Twitter data. Proceedings of the 23rd International Conference on World Wide Web, pp. 439–444.

Zagheni, E., I. Weber and K. Gummadi

2017 Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 43(4):721–734.