Openness to Migrate Internationally for a Job: Evidence from LinkedIn Data in Europe

Daniela Perrotta,[1]*Sarah C. Johnson,[1]*Tom Theile,[1] André Grow,[1] Helga de Valk,[2,3] Emilio Zagheni[1]

[1]Max Planck Institute for Demographic Research
[2]Netherlands Interdisciplinary Demographic Institute
[3]University of Groningen
perrotta@demogr.mpg.de, johnson@demogr.mpg.de, theile@demogr.mpg.de, grow@demogr.mpg.de, valk@nidi.nl, zagheni@demogr.mpg.de

Abstract

Understanding the factors that explain why people move – or stay – and where they go, is a central goal of migration research. This article improves our understanding of migration aspirations of professionals in Europe by leveraging a previously untapped data source: aggregate-level information on LinkedIn users open to work-related international relocation, accessed through the LinkedIn Recruiter platform. We collected data at regular intervals from Oct. 2020 to Sept. 2021. First, we offer descriptive statistics of proxies for migration aspirations (or lack thereof) for millions of Linkedin users in Europe. Then we approach our questions using a standardization technique, based on gravity models of migration, in order to account for a number of biases in the data, including uneven use of LinkedIn across countries. We found that, in absolute terms, countries in Northern and Western Europe are the most attractive ones when considering LinkedIn users open to work-related relocation (about 60%), followed by Southern Europe (about 40%) and Eastern Europe (30%). We also observed substantial heterogeneity in directionality of aspirations: for example, roughly 20% of LinkedIn users would relocate from Western to Northern Europe, while less than 10% would relocate from Northern to Western Europe. After accounting for differences in population density, geographic and linguistic distances, as well as internet and LinkedIn penetration, we observed that, in relative terms, Southern Europe appears to be a highly desired destination for professionals, indicating that there is potential for changing patterns in actual flows, should more opportunities for professionals arise in Southern Europe.

Broader Perspective and Significance

International migration is an important phenomenon with measurable demographic consequences. It has therefore become of paramount importance for policymakers in Europe and globally to understand the key drivers that shape migration potentials across countries for migration preparedness, for planning purposes, and for policy implementation. Yet, a lack of data and of appropriate methods to extract reliable information from often noisy or deficient data sources limits our ability to measure and predict migration and mobility.

In this article, we unveil a novel dataset we collected via the LinkedIn Recruiter platform in order to reduce this data gap and to improve our understanding of the demand for migration of professionals in Europe. Since the LinkedIn Recruiter platform is designed for professional recruiting purposes and not for use in demographic and social research, data pre-preprocessing was an important first step to ensure data quality and to avoid introducing biases in our analyses. It is important to note that we collected only anonymous, aggregate-level data, from which identification of individuals is impossible.

To the best of our knowledge, our study provides the first most comprehensive and rigorous cross-national and comparative data on international migration aspirations from online social media data. We believe that, beyond the substantive results of this article, and the academic value of this research, the data and statistical resources that we present hold the potential to become important tools for guiding decision-making processes of policy makers in Europe and beyond.

Introduction

International migration is an important phenomenon with measurable demographic consequences. Much migration research, though, is plagued by inadequate data and leaves much to be explained as to why people choose to move – or stay – and $where$ they move to (Willekens et al. 2016). Recent theoretical advances emphasize a "two-step" approach to the study of migration, called the "aspiration-(cap)ability" framework, in which the decision-making process prior to migration is separated from the migration event (or lack thereof) (Carling and Schewel 2018; De Haas 2021). In addition to this theoretical framework, the number of surveys that focus on people who have a migration "aspiration" has increased over the last decade. This has enabled researchers to improve on various dimensions of the aspiration-(cap)ability framework (Aslany et al. 2021).

While traditional surveys of migration aspirations are of great use for theory development, they typically have several shortcomings. First, the instruments that inquire about a specific dimension of a person's migration aspirations are often tailored to a single research question and therefore query only one dimension of a person's migration aspirations. Second, they tend to be limited in time and space, as they are

[*]These authors contributed equally.

often cross-sectional and geographically restricted (a recent review found only 7% of them are multi-regional (Carling and Mjelva 2021)). Third, while potentially representative of the general population, they often lack inter-survey comparability (due to diversity across survey instruments) and suffer from acquiescence bias (as many questionnaires ask about preferences for leaving but not staying).

To address some of the limitations of existing survey data and to provide a perspective that complements the growing literature in the field, we offer new analyses that rely on a novel – and as of yet untapped – data source: aggregate-level information on LinkedIn users open to job-related relocation, as obtained from the LinkedIn Recruiter platform. Compared to traditional survey data, the LinkedIn data that we employ here to study openness to migration (1) are continuously available, (2) have consistently defined variables across 24 different languages, and (3) provide a global snapshot of openness to migration as recent as the latest update to a person's LinkedIn profile. Furthermore, rather than directly asking a person about their migration preferences, we capture routine job-seeking behaviors, thereby avoiding reactive responses to a survey instrument.

The main aims of our work are to identify the utility and limitations of this novel LinkedIn dataset for studying openness to international migration. We use the term "openness" to migration to more accurately reflect the terminology used on LinkedIn, and we refer to the concept of migration "aspirations" used in the literature as part of an imagined future where migration is a possibility. Hence, we study a more general receptiveness to the idea of migration, rather than a concrete desire. In what follows, first we describe the data set and data collection process. Then we present descriptive results on inter-regional trends. Finally, we compare the relative attractiveness of certain countries by means of a gravity-type model, that controls for a set of geographic and linguistic factors that may generally affect openness to move from one country to another (Cohen et al. 2008). In our analyses, the focus is on Europe. The reasons for this are two-fold. First, there is a large number of LinkedIn users who are open to move to and within Europe. Second, Europe offers an interesting case study, due to the establishment of free movement of European Union (EU) nationals across borders. In this context, we identify pairs of countries where openness to migration is different from what we would expect based on the geographic and linguistic factors alone, which we control for in our gravity model. As such, our work offers a potential indicator of future migration flows within Europe.

Background

Migration is a complex, heterogeneous and selective process. It is also a driving force of economic, demographic, social and political change. Periods of crises have reminded us of the importance of migration and mobility for our societies, both in the short- and in the long-term.

Migration has become a top priority for policymakers in Europe and globally, as our societies increasingly face the challenges of managing vital migration flows and integrating migrants in the context of below-replacement fertility levels, slow population aging and sudden crises or shocks. The United Nations 2030 Sustainable Development Goals include an international commitment to "facilitate orderly, safe, regular, and responsible migration and mobility of people, including through the implementation of planned and well-managed migration policies" (target 10.7). This is consistent with the Global Compact for Safe, Orderly and Regular Migration, which was adopted by the majority of UN Member States on December 10, 2018. The Global Compact marked the first time that Heads of State and Government came together, at the global level, under the auspices of the United Nations, to address the issue of migration in a comprehensive and collaborative way. Policymakers have also increasingly recognized the role of attracting "the best and the brightest" (Kapur and McHale 2005) in the global competition for talent, in order to favor economic growth and productivity.

Despite the importance of migration to understand social and economic change and the need for deep comprehension of migration processes in order to devise sensible policy interventions, our ability to measure and predict migration and mobility remains limited. One of the main factors that hinders the advancement of our understanding of migration in human populations and the further development of theoretical frameworks is a lack of data and of appropriate methods to extract reliable information from often noisy or deficient data sources.

Early migration theories attempted to provide a generalized understanding of migration processes. A classic example of these efforts is Ravenstein's Laws of Migration (Ravenstein 1889) that, among others, introduced the concepts of push and pull factors in migration. Contributions to the field in the second half of the 20th century also attempted to provide a rather comprehensive account of migration dynamics. They include Lee's theory of migration (Lee 1966), the theory of mobility transition (Zelinsky 1971), neo-classical migration theory (Harris and Todaro 1970), dual labor-market theory (Piore 1979), new economics of labor migration theory (Stark 1978), and the cumulative causation theory (Massey et al. 1993). More recently, with some exceptions, migration research has moved away from comprehensive theories and has focused on more specialized studies (De Haas 2021). This has partially contributed to the consolidation of a number of divides. They include the divide between international and internal migration (Skeldon 2006), the divide between micro and macro processes (Bijak 2006), as well the divides related to the temporal scale for the analysis of long-term (or permanent) migration versus short-term (or onward) mobility (Fiorio et al. 2021). These divides cut across broad lines of literature, including the ones that investigate the relationship between migration and development, the role of social networks for the initiation and continuation of migration processes, and the relationship between long-term trends versus sudden shocks or discontinuities in shaping migration corridors (Massey et al. 1993).

To advance our theoretical understanding of migration, to better understand and explain migration patterns, as well as to improve our predictive capacity, requires combining all available data on migration flows with theoretical knowledge about migration. Quantitative information on inflows and outflows of migrants is fundamental to understand drivers and consequences of migration. Inflows and outflows are part of the so-called basic demographic equation, on which all more complex demographic models are built. However, in practice, due to lack of data or data quality issues, inflows and outflows are often simplified in demographic accounting and modelling.

In the European context, the most comprehensive collection of data on migration flows is maintained by Eurostat, which publishes official statistics that it receives from national statistical offices. These data collections rely on censuses, surveys and national administrative sources, which include population registers, border data collection systems and visas, residence permits, and/or work permits. The quality of the data depends on the country's migration/registration procedures, the legal incentives for registering the migration event, and the methodologies used by national statistical offices to measure migration. While these data are comparatively better than in other parts of the world, they suffer from inadequacies that drastically limit their usefulness. For example, the flows from Poland to Germany vary by more than one order of magnitude depending on whether they are reported by Poland or Germany, as both countries use different definitions and data collection systems. Eurostat provides flows from Poland to Germany and from Germany to Poland only for limited years, and since 2009 these data are not available. This availability is further restricted when such bilateral data are disaggregated by age and sex. This is not an isolated case. It is the result of some key limitations related to administrative data and lack of a comprehensive view that combines multiple sources to assess migration in Europe. Moreover, data become available only with a substantial delay, making it very difficult to gauge present trends. In some cases, data are not published for several years.

It is widely recognized that migration data in Europe differ widely across countries and that combining data sources is key to improve our understanding of migration flows (Willekens 1994). Over the last three decades a first line of research has developed statistical models to combine quantitative and qualitative data with the goal of producing synthetic databases of estimates of migration flows (Willekens et al. 2019). Key methodological innovations included the use of Bayesian statistical models (Raymer et al. 2013), the combination of administrative data and survey data (Wiśniowski 2017), as well as indirect approaches to estimate flows from migrant stocks (Abel and Sander 2014; Azose and Raftery 2019). This important line of research has developed practices to assess data quality and to address data imperfections through statistical modelling. Expert opinions as well as qualitative information, assumptions or theoretical considerations are incorporated into the models via the use of Bayesian methods and appropriate approaches for elicitation of information, like the Delphi method.

In the last decade, a second line of research on migration estimation has emerged as a result of the so-called "data revolution" and the digitalization of life. Digital trace data, including Web and social media data, have opened up new opportunities for studying and understanding migration. Following a pioneering study using e-mail data and IP geolocation to estimate international migration flows (Zagheni and Weber 2012), new approaches to study migration and mobility that rely on social media data have emerged. They include the development of methods to use Twitter data to assess the relationship between short-term mobility and long-term relocations (Fiorio et al. 2017, 2021), or to infer international migration (Zagheni et al. 2014). Historical information from Wikipedia have been used to understand the role of migrations and international collaborations in the context of innovation and cultural development (Lucchini, Tonelli, and Lepri 2019). Facebook data have been used to assess the impact of natural disasters on short-term mobility (Alexander, Polimis, and Zagheni 2019), and, more broadly, to quantify international mobility (Zagheni, Weber, and Gummadi 2017; Spyratos et al. 2019). In addition to that, Facebook data have been combined with existing survey data within a Bayesian hierarchical model to produce estimates of migration stocks in the United States (Alexander, Polimis, and Zagheni 2020).

As part of this 'data revolution', LinkedIn data has been used to measure migration flows of professionals to the United States (State et al. 2014). Our article builds on, and advances, this broad line of literature that aims at complementing traditional sources with social media data for the study of migration. More specifically, it goes one step further by leveraging an untapped data source for the study of migration (LinkedIn Recruiter data) and by addressing a new and relevant issue, which is the timely assessment of the potential demand for international relocation. Measuring openness to migration is the first step towards assessing the link between aspirations and actual outcomes. As far as the authors know, there have not been previous attempts at using social media data to estimate proxies for migration aspirations. We hope that this article would stimulate productive collaborations across scientific communities to fully leverage the potential of social media for migration research.

LinkedIn Dataset

LinkedIn is a professional networking site of nearly 800 million users. Its main purpose is to connect professionals with each other and to new job opportunities. Within LinkedIn, the so-called Recruiter platform enables recruiters from subscribing companies to identify potential job candidates through users' profile attributes, such as industry, educational attainment, and years of experience. To avoid gender- and age-based discrimination in hiring practices, the LinkedIn Recruiter platform does not allow us to directly search for these attributes. Of particular interest for this study is the ability to search LinkedIn via the Recruiter platform for users open to job-related relocation. This latter refers to those users who have indicated in their profiles that they are open to finding a new job, and have listed one or more prospective job locations that differ from where they are currently located. We refer to this group of users as those who are open to job-related international migration, thus referring to the concept of migration aspirations in the literature, but more accurately reflecting the definition used on LinkedIn.

Data Collection

We collected data from the LinkedIn Recruiter platform regularly every two weeks, from 2020-10-08 to 2021-09-06, for a total of 25 data points. At each iteration, we collected both the aggregate number of LinkedIn users and of LinkedIn users open to work-related relocation internationally. The data collection process involves a separate search query for each destination country and returns the top 75 current user locations (ranked by number of users). As an illustrative example, a single search would collect the number of LinkedIn users open to relocating to Germany (but currently not located in Germany) stratified by their current location. Note that while the spatial resolution of the prospective job location can be specified in the input field as desired (e.g. country), the spatial resolution of the current job locations may vary from metropolitan area up to the country level and cannot be selected. Typically the platform returns a sorted list of locations with the largest populations of LinkedIn users.

It is important to note that we collect only anonymous, aggregate-level data, from which identification of individuals is impossible. The data were collected purely for scientific purposes using the LinkedIn Recruiter Platform, accessible at the following URL: https://www.linkedin.com/talent/. More specifically, the data collection was performed using the API provided by the LinkedIn Recruiter Platform.

Data Pre-processing

In this work, we focus on Europe and consider a spatial resolution at the country level, thus dropping any other lower spatial resolution (e.g. metropolitan area). Note that some countries may appear in the dataset with different names (e.g. Czech Republic and Czechia). However, this is only a naming issue: absolute values are unique and consistent when looking at common country pairs (e.g. number of users open to relocating from Czech Republic or Czechia to Germany).

In the European context, Liechtenstein is missing and we drop data for Cyprus. The latter in fact appears in the dataset also as Cyprus UN Neutral Zone, but with unexpectedly higher number of users open to relocation (over three orders of magnitude compared to Cyprus). This is likely due to an incorrect country selection on LinkedIn.

Due to the variability in the spatial resolution and due to the truncation at the top 75 current job locations, the resulting data collection varies considerably across countries and query dates. Figure 1 shows the variability in the number of times each European country appears in a pair of countries as current versus prospective job location, across query dates. It is evident that this bias affects particularly those countries with fewer LinkedIn users which are likely excluded from the resulting top 75 current job locations. As an example, Iceland, which has the smallest number of LinkedIn users in Europe, never showed up as current job locations due to this top 75 cutoff. On the other hand, countries with higher numbers of LinkedIn users, such as the United Kingdom, France, Italy, Spain, and Germany, are substantially more likely to appear in the search. To address this bias and to avoid any imbalance due to the data collection scheme, in this analysis we consider only those country pairs

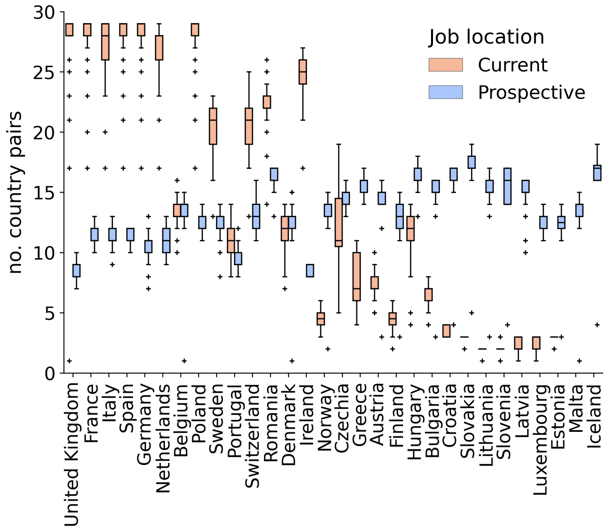

Figure 1: Variability in the size of country pairs in which each of the 30 European countries appears as current versus prospective job location, across query dates. Countries are sorted based on the total number of LinkedIn users, ranging from nearly 30 million users in the United Kingdom to approximately 150 thousand users in Iceland.

.

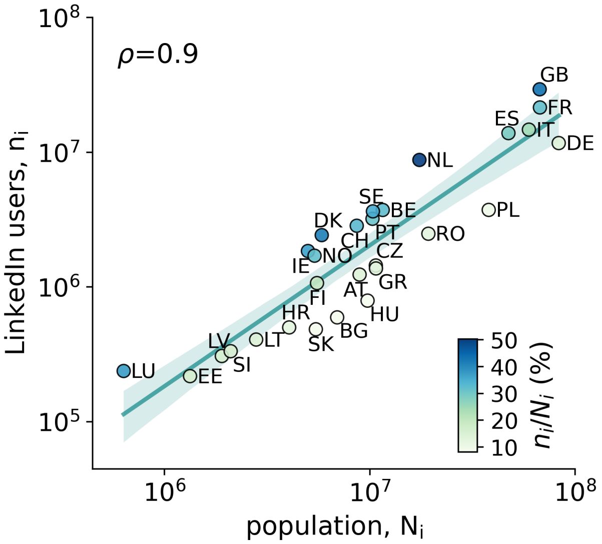

Figure 2: Relationship between the LinkedIn population sample (averaged across query dates) and the general population by European country (reported as ISO 3166-1 alpha-2 country codes). The colour code refers to the population sampling ratio n_i/N_i. The Spearman correlation coefficient between the two datasets is $\rho = 0.9$, $p < 10^{-11}$.

.

having both bilateral “flows”, w_{ij} and w_{ji}, of LinkedIn users open to job-related international relocation from country i to country j and vice versa. This choice drastically reduces the size of our dataset (by roughly 45%), but ensures data comparability across the countries that we prioritize for this study. Please note that, for the remainder of the article, we refer to the number of LinkedIn users open to job-related relocation from country i to country j (w_{ij}) as “flows” from country i to country j. This is for simplicity of language only, even though we do not observe actual flows. Instead we observe openness to relocation, or potential flows.

Due to the variability in the dataset, here we use median country-level flows across all dates of data collection. We employ a standard weighting approach to normalize flows w_{ij} based on the population sampling ratio n_i/N_i, where n_i is the LinkedIn population sample and N_i is the general population in country i (population data from (The World Bank 2020a)). This way we correct for potential biases due to under- or over-sampling the population, although population samples are already in good agreement, as shown in Figure 2 (Spearman’s $\rho = 0.9$, $p < 10^{-11}$).

Descriptive Analysis

The resulting dataset consists of a total of 28 European countries, 25 time data points and 5,222 unique queries of LinkedIn users open to job-related international relocation across countries. Figure 3 shows the openness to job-related international relocation in Europe in form of origin-destination matrix. The heatmap shows that the flows w_{ij} and the number of links ij tend to decrease with smaller population size. The matrix is not symmetrical, thus revealing those countries that may act as a source or sink for future potential migration. This is more evident in Figure 4 that shows the relationship between the inflows $w^{in} = \sum_i w_{ij}$ and the outflows $w^{out} = \sum_j w_{ij}$ by country. Specifically, for a given country i, if $w^{in} > w^{out}$ it means that, overall, the number of LinkedIn users open to entering the country is greater than the number of LinkedIn users open to leaving the country. This is the case, for example, for Luxembourg (LU), Switzerland (CH), and the Netherlands (NL). On the other hand, when $w^{in} < w^{out}$ it means that the country would potentially lose more migrants than those gained. This is the case, for example, for Poland (PL), Greece (GR), and Romania (RO). Other countries, such as Portugal (PT) and Estonia (EE), have instead roughly the same inflows and outflows.

Figure 5 provides an overview of the potential mobility patterns within subregions of Europe based on the United Nations geoscheme. In more detail, the figure shows the proportion of LinkedIn users open to relocation between two countries, aggregated to European regions. Here proportions are scaled within each region’s total prospective turnover, that is, the total number of users currently located and/or open to relocating to that region. We observe, for example, that roughly 60% of LinkedIn users open to relocation would relocate to Western and Northern Europe, while about 40% would relocate to Southern Europe and only 30% to Eastern Europe.

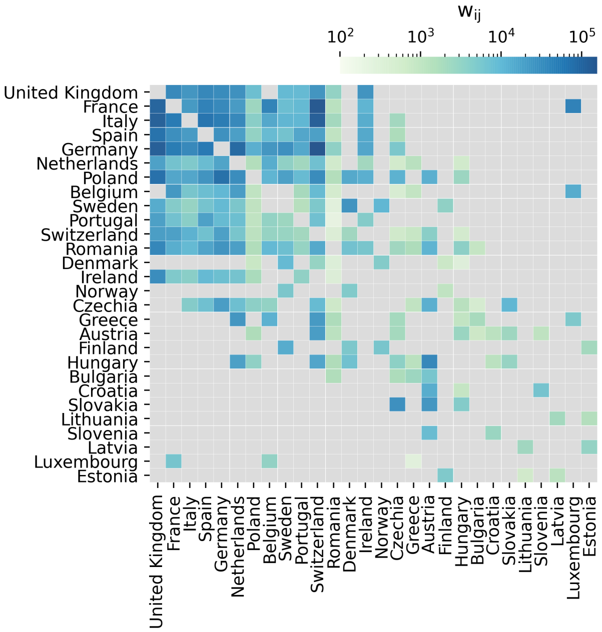

Figure 3: Origin-destination matrix of LinkedIn users open to job-related international migration from country i (on the y-axis) to country j (on the x-axis). The colour code represents the normalized flows w_{ij} (grey indicates no data). Countries are sorted according to LinkedIn population size.

.

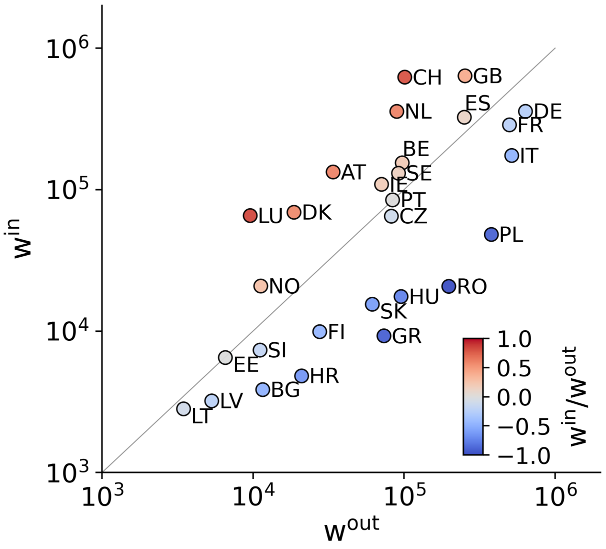

Figure 4: Relationship between the inflows w^{in} and outflows w^{out} by country. The colour code refers to the ratio w^{in}/w^{out} (log scale) and separates migrant-receiving versus migrant-giving countries. The solid line $x = y$ is a guide to the eye.

.

The fact that we have been collecting data repeatedly makes it possible to explore potential changes in migration

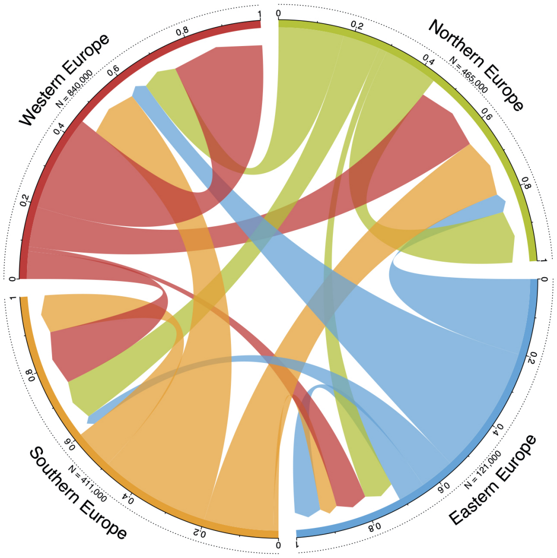

Figure 5: Chord diagram showing the number of LinkedIn users open to relocation between two countries, aggregated to European regions, based on the United Nations geoscheme. Direction is indicated by the arrowhead and size by the width at the base of the arrow. Proportions have been scaled within each region, with the total rounded to the nearest thousand. Each region's N represents the total prospective turnover, i.e. all those currently located in the region and open to relocating plus those who are open to relocating to the region from elsewhere. As an illustrative example, nearly 20% of those in Northern Europe or open to relocating to Northern Europe would like to move to Western Europe.

.

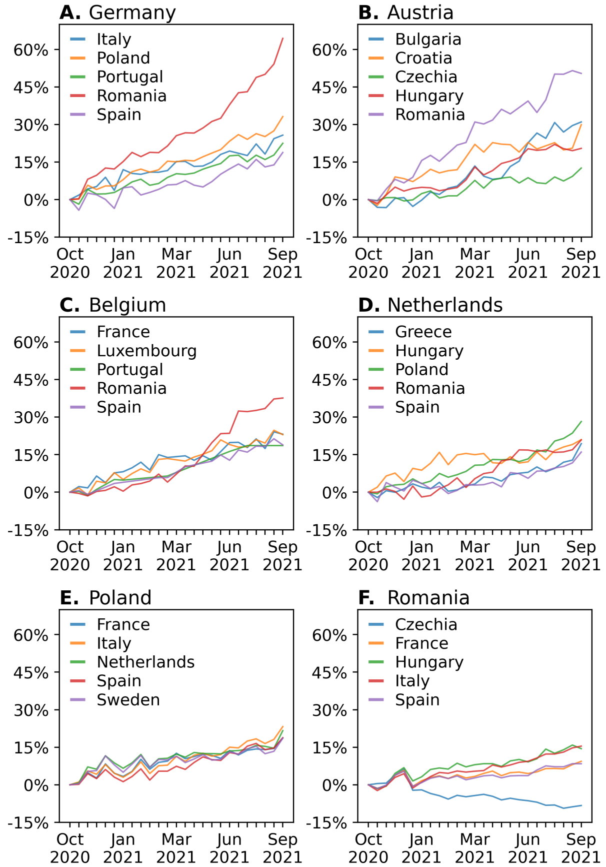

Figure 6: Percent change in the flows w_{ij} from the origin country, A) Germany, B) Austria, C) Belgium, D) the Netherlands, E) Poland, and F) Romania, to the corresponding top 5 destination countries having highest variation over time. The percent change is calculated by considering the initial value on the first date of data collection.

.

desires over time. We computed the percent change in the normalized flows w_{ij} by comparing each data point to the initial value on the first date of data collection on 2020-10-08. Figure 6 shows this variation for a few interesting origin countries (i.e. Germany, Austria, Belgium, the Netherlands, Poland, and Romania) to the corresponding top 5 destination countries having highest variation over time. We observe, for example, that Romania has the highest increase in migration desires from LinkedIn users located in Germany (over 60%), Austria (over 50%) and Belgium (over 30%), potentially a sign for desired return migration. On the contrary, when considering Romania as the origin country, this has less variation over time in terms of migration desires to new countries. Interestingly, most temporal trends are increasing, except for a few cases, such as from Romania to Czechia, as shown in Figure 6F.

Modelling the Openness to Migration

Interpretation of the numbers in Figure 5 makes inter-country comparisons challenging, since differences could be attributed to, say, the difference in LinkedIn users between two countries. We therefore use a modeling approach as a form of standardization to assess the relative attractiveness of prospective job-related relocation between countries. In other words, here we are interested in identifying those countries that are more (or less) attractive in terms of future potential migration as regulated by a number of factors that may generally affect the openness to move from one country to another. For this, we largely follow the state-of-the-art gravity-type model proposed by (Cohen et al. 2008). In general, the gravity model assumes that the number of people moving between two locations i and j is proportional to the population size in i and j, scaled by their distance d_{ij}. Cohen and co-authors proposed an adaptation of the gravity model based only on geographic and demographic independent variables to project international migrations across countries (Cohen et al. 2008). Following the same methodology, here we estimate the expected number of users N_{ij} open to relocating from their current country i to a prospective destination country j using the following equation:

$$
\begin{aligned}
\log(N_{ij}) = \beta_0 + \beta_1 \log(P_i) + \beta_2 \log(P_j) + \\
\beta_3 \log(D_{ij}) + \beta_4 \log(A_i) + \beta_5 \log(A_j) + \\
\beta_6 \log(I_i) + \beta_7 \log(I_j) + \beta_8 \log(L_{ij}) + \epsilon_{ij}
\end{aligned}
\quad (1)
$$

where $\log$ refers to $\log_{10}$, P_i and P_j are the median numbers of LinkedIn users in countries i and j, D_{ij} is the their distance, A_i and A_j are the area of countries in km^2 excluding area under inland waters and coastal waters in 2017 (FAO 2020), I_i and I_j are the proportions of the population using the internet in 2019 (The World Bank 2020b), and L_{ij} is the probability that two people in countries i and j understand a common language (Melitz and Toubal 2012). Distance, here, is the average bilateral population-weighted geodesic distance between the most populous cities of two countries, and is appropriate for migration-related analyses as it accounts for non-uniform population density distributions in different countries.[1]

The linear model is fitted with the dependent variable $log(N_{ij})$ as the median country-level bilateral flows collected from LinkedIn. Starting from the standard gravity law with the three basic independent variables (i.e numbers of LinkedIn users P_i and P_j, and distance D_{ij}), we employ a stepwise algorithm search approach with Bayesian information criterion to obtain the best model. The model in Eq. 1 is the final best model selected. The regression achieves the following values in terms of measure of fit: R^2 = 0.827 and adjusted R^2 = 0.821. Table **??** reports the estimated values and statistics for all the free parameters in Eq. 1. Note that comparably similar fit results are obtained when omitting the area of the origin and/or destination country.

Descriptive Bivariate Relationships

The number of LinkedIn users open to job-related relocation increases with increasing LinekdIn user size of both origin (r=0.567) and destination (r=0.626) country, increasing area of origin country (r=0.332), increasing internet penetration in the destination country (r=0.561), and increasing probability of understanding a common language (r=0.375). Correlations with the population-weighted distance between countries, area of the destination country, and internet penetration of the current country, are insubstantial. Log LinkedIn user size and log area were correlated for origins (r=0.488) and destinations (r=0.494). Because these were the two highest-magnitude correlations, collinearity was not a problem in fitting the model. Log of the population-weighted distance is negatively correlated with the probability of understanding a common language (r=-0.417). This is not surprising in the European context, where most countries speaking a common language also share a border.

Coefficient	Estimate	Standard Error
(Intercept)	-6.337***	1.832
$\log(P_i)$	0.528***	0.048
$\log(P_j)$	0.67***	0.049
$\log(D_{ij})$	-0.532***	0.076
$\log(A_i)$	0.128*	0.052
$\log(A_j)$	-0.107*	0.052
$\log(I_i)$	-2.532***	0.620
$\log(I_j)$	4.307***	0.620
$\log(L_{ij})$	0.603***	0.110

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Resulting coefficients and statistics of the free parameters in the gravity model of Eq. 1 as obtained by applying a multivariate linear regression to the LinkedIn data on job-related relocation among countries in Europe. All p-values are statistically significant at the 0.05 level.

Model Results

We used the gravity model as a form of standardization to assess the relative attractiveness of prospective job-related relocation between European countries based on the population density of LinkedIn users between countries, their geographic and linguistic distances, and internet usage differences. We therefore compared the observed and predicted values of LinkedIn users open to job-related relocation across countries in Europe in order to identify those countries that are more (or less) attractive for relocation for a new job, compared to what would be expected from the model predictions alone. For this, we computed the percentage error as a measure of the discrepancy between the predicted and observed values. Figure 7A shows the percentage error divided into quintiles, ranging from observed values much lower than predicted (in blue) to observed values much higher than predicted (in red). As an illustrative example, the United Kingdom turns out to be a much more attractive location for a new job for individuals in Italy and Spain than predicted by the model, while it is less attractive (relatively speaking, with reference to baseline model predictions) for individuals in Germany, the Netherlands, and Sweden. Some countries are relatively more attractive than predicted. This is the case for example of Italy, Spain, Germany, Portugal, Switzerland, and Austria. Conversely, other countries, such as the Netherlands, Belgium, Poland, Sweden, and Romania, are much less attractive according to the observed data than to what is predicted by the model. Furthermore, we observe a number of country pairs exhibiting relatively similar high reciprocal attractiveness, such as Italy and United Kingdom,

[1] The geodesic distance is calculated using the following expression:

$$D_{IJ} = \frac{\sum_i^I w_i \sum_j^J w_j d_{ij}}{\sum_i^I w_i \sum_j^J w_j}$$

where D_{IJ} is a distance between two countries I and J; i is a city of country I and j is a city of country J; w is the city population taken from $world.cities$ dataset (Becker and Deckmyn 2018); and d_{ij} is the geodesic distance between two cities. The calculation of the d relies on the commonly used World Geodetic System 84 reference ellipsoid (Karney 2013; Hijmans 2019). There are, at most, 50 cities included for each country; if a country has fewer than 50 cities, all available cities are included.

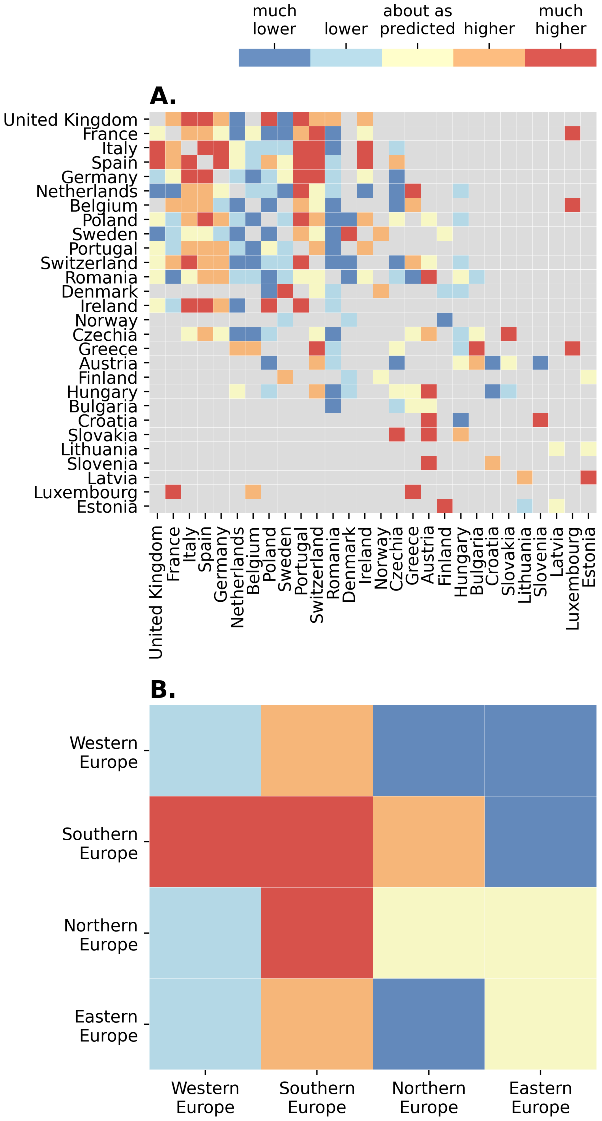

Figure 7: Percentage error between predicted and observed values of LinkedIn users open to relocate between origin (y-axis) and prospective destination (x-axis) countries (A) and European regions (B). Countries are sorted according to LinkedIn population size. Each cell is colored based on the quintile of the percentage error, with observed values much lower than predicted in blue and observed values much higher than predicted in red (grey indicates no data).

.

Sweden and Portugal, and France and Luxembourg.

Figure 7B shows the quintiles of percentage error when we aggregate the potential flows at the level of European regions. Here we observe that Southern Europe is the most attractive job location for individuals from all European regions. This means that, while, in absolute terms, the number of LinkedIn users open to relocate to Southern Europe is relatively small, after we account for differentials in LinkedIn and internet penetration rates, and other key explanatory factors included in the gravity model, Southern Europe appears to be a highly desired destination in relative terms. By contrast, Eastern Europe is much less attractive for individuals in Western and Southern Europe. Western Europe results to be much more attractive solely for individuals in Southern Europe. Note that in this spatial aggregation not all European regions are fully represented as the network is sparse.

Discussion

The aim of this work was to identify the utility and limitations of a novel LinkedIn dataset as a proxy for studying migration aspirations of professionals in Europe. For this, we leveraged a previously untapped data source, via the LinkedIn Recruiter platform, and collected aggregate-level data on LinkedIn users' openness to work-related international relocation. These represent potential future migration flows. The resulting dataset consisted of unique bilateral flows across 28 European countries during the time period between October 2020 and September 2021. We used these data to build an origin-destination matrix of cross-national openness to relocation. This allowed us to identify countries as more (or less) attractive in terms of future potential migration. As a form of standardization, we then employed a gravity-type model to assess the relative attractiveness of countries. For this, we largely followed the state-of-the-art modeling approach proposed by (Cohen et al. 2008) and modelled migration flows as regulated by a number of factors that may generally affect openness to move from one country to another, namely, LinkedIn population density, geographic and linguistic distances, and internet penetration differences. Our model thus offers a baseline to which we compare the observed LinkedIn values to assess the relative importance of these factors in shaping potential future flows of professionals in Europe and the relative attractiveness across European countries.

Our findings show that countries in Northern and Western Europe are the most attractive ones for LinkedIn users open to work-related relocation, with about 60% of potential incoming flows, followed by Southern Europe with about 40% and Eastern Europe with only 30%. The composition of the flows in terms of origin and destination locations is very diverse. For example, looking at the potential mutual exchange of job seekers between Western and Northern Europe, roughly 20% would relocate from Western to Northern Europe, while less than 10% would relocate from Northern to Western Europe. By contrast, much more LinkedIn users from Southern Europe would be open to relocate to Western Europe (about 30%) than to Northern Europe (about 20%). Additionally, we found large differences in the openness of LinkedIn job seekers to relocate to a country within the same European region where they are currently located. This range from over 25% for Western Europe, 15% for Northern Europe, over 10% for Southern Europe, and less than 10% for Eastern Europe.

A gravity-type model accounts for a large component of the variability in the observations. We consider the predictions of the model as baselines for what we would expect to see in the data, after controlling for key variables like LinkedIn and internet penetration rates. Comparing the actual LinkedIn observations with the respective predicted values confirms that Eastern Europe remains a relatively unattractive destination. On the other hand, Southern Europe appears to be more attractive than what it seemed based only on descriptive trends. It is unclear what drives this demand. We may speculate that this could be partially the result of labor migrants from Southern Europe, living in other parts of Europe, who are open to returning to their home countries if opportunities arise. Future research, potentially including a combination of passively-collected data and survey data, is needed to shed more light into these patterns.

Limitations and Challenges

The data that we presented here fill a niche, by offering large-scale, repeated, and non-reactive measurements of people's openness to relocate to other countries during job search. These advantages notwithstanding, these data also come with some limitations and challenges. First, as any online social media platform, LinkedIn's user population is not representative of the general population, and our sample could be potentially biased due to self-selection and non-representativeness. Attention to what types of data are used and who is represented in the data are indeed critical to avoid limiting the validity of conclusions drawn. In Figure 2, we showed that the LinkedIn population sample is in good agreement with the general population in each country. Additionally, in our analyses we used normalized flows based on the population sampling ratio in order to correct for potential differences between the LinkedIn populations and the general populations. We also went further by assessing the observations against benchmark predictions from a gravity model in order to filter out biases related to differentials in variables like LinkedIn and internet penetration rates across countries. However, accounting for all possible biases, including those related to unobservable characteristics and desires, is beyond the scope of this article.

The second set of limitations pertains to the way the recruiting platform is designed, inevitably affecting our data collection and potentially hindering scalability and/or reproducibility of this methodology as the data source might change at LinkedIn's discretion. First of all, the top 75 cutoff in the origin locations that are returned for each target destination location represents another source of bias toward countries with higher LinkedIn populations. We addressed this issue by considering only those country pairs that have bilateral "flows" of job-related openness to migration, thus drastically reducing the size of our dataset and inevitably losing information pertaining to smaller countries. Additionally, the LinkedIn Recruiter platform does not allow us to directly search for users by age, gender, or nationality, as a necessary precaution to avoid specific discrimination in hiring practices. This represents a limitation in our analysis, as previous studies have shown age- and gender-specific patterns in high-skilled migration (d'Aiglepierre et al. 2020; Kashyap and Verkroost 2021), and hinders the potential use of ad-hoc post-stratification weights that would be relevant to approximate a representative sample of the general population, at least in the central demographic variables.

Lastly, although the data collected from the LinkedIn Recruiter platform allow us to identify potential future migration flows across countries, these data alone are not enough to link those who expressed a migration desire and those who actually migrate. However, previous studies showed that the number of people planning to migrate is a good predictor of actual flows of people and are critical to help develop migration scenarios and forecasting models (Tjaden, Auer, and Laczko 2019; Laczko, Tjaden, and Auer 2017).

Conclusions and Future Work

Our study contributes to improving our understanding of migration aspirations of professionals in Europe. To the best of our knowledge, our study represents the first initial step to characterize international migration aspirations from online social media data and contributes to reducing the data gap in migration potentials. We showed the utility and limitations of leveraging a novel LinkedIn dataset for studying job-related openness to international migration and how the relative attractiveness between European countries can be quantified by means of a gravity-type model. More recent efforts have been made in the modeling of migration flows using various adaptations of the gravity and radiation models (Simini et al. 2021; Lucchini, Tonelli, and Lepri 2019; Beiró et al. 2016). Comparing the outputs obtained from different modeling approaches is beyond the scope of this article, but we consider it an important aspect, that we will explore in future work, to assess their relative predictive power. Future work will also delve into other additional country-level factors driving attractiveness amongst job-seekers and shaping cross-national migration flows, including the socio-demographic and educational features that we can collect from the recruiting tool (e.g., industry category, years of experience, skills). While here we focused on specific factors related to geographic and linguistic distances, in future work we will explore other factors that may be relevant in the decision-making process when considering to migrate to another country, such as socio-economic or environmental factors. At the same time, more in-depth validations of the LinkedIn data against survey data is needed in order to assess the added value of this tool for use in demographic research. Much of the recent literature investigating potential migration aspirations is based on the Gallup World Poll that every year conducts nationally representative surveys in over 160 countries and provides an indication of migration intentions and preparations (Migali and Scipioni 2018). Micro data from the Gallup World Poll are proprietary data and highly expensive to purchase. We could not access this data source, and the survey results from Gallup are not immediately comparable with the ones of LinkedIn data. However, we expect that future research could focus on combining the two sources in order to improve our understanding of trends and nuances for migration aspirations.

References

Abel, G. J.; and Sander, N. 2014. Quantifying global international migration flows. *Science*, 343(6178): 1520–1522.

Alexander, M.; Polimis, K.; and Zagheni, E. 2019. The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data. *Population and Development Review*, 617–630.

Alexander, M.; Polimis, K.; and Zagheni, E. 2020. Combining social media and survey data to nowcast migrant stocks in the United States. *Population Research and Policy Review*, 1–28.

Aslany, M.; Carling, J.; Mjelva, M. B.; and Sommerfelt, T. 2021. Systematic review of determinants of migration aspirations. *Changes*, 1: 18.

Azose, J. J.; and Raftery, A. E. 2019. Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences*, 116(1): 116–122.

Becker, R. A.; and Deckmyn, A. 2018. *Maps: Draw Geographical Maps*.

Beiró, M. G.; Panisson, A.; Tizzoni, M.; and Cattuto, C. 2016. Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Science*, 5: 1–15.

Bijak, J. 2006. Forecasting international migration: Selected theories, models, and methods. Citeseer.

Carling, J.; and Mjelva, M. B. 2021. Survey instruments and survey data on migration aspirations. *Changes*, 1: 18.

Carling, J.; and Schewel, K. 2018. Revisiting aspiration and ability in international migration. *Journal of Ethnic and Migration Studies*, 44(6): 945–963.

Cohen, J. E.; Roig, M.; Reuman, D. C.; and GoGwilt, C. 2008. International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences*, 105(40): 15269–15274.

De Haas, H. 2021. A theory of migration: the aspirations-capabilities framework. *Comparative Migration Studies*, 9(1): 1–35.

d'Aiglepierre, R.; David, A.; Levionnois, C.; Spielvogel, G.; Tuccio, M.; and Vickstrom, E. 2020. A global profile of emigrants to OECD countries: Younger and more skilled migrants from more diverse countries. *OECD Social, Employment and Migration Working Papers*, (239).

FAO. 2020. FAOSTAT Land Use domain. http://www.fao.org/faostat/en/#data/RL. Accessed: 2021-12-01.

Fiorio, L.; Abel, G.; Cai, J.; Zagheni, E.; Weber, I.; and Vinué, G. 2017. Using Twitter data to estimate the relationship between short-term mobility and long-term migration. In *Proceedings of the 2017 ACM on web science conference*, 103–110.

Fiorio, L.; Zagheni, E.; Abel, G.; Hill, J.; Pestre, G.; Letouzé, E.; and Cai, J. 2021. Analyzing the effect of time in migration measurement using georeferenced digital trace data. *Demography*, 58(1): 51–74.

Harris, J. R.; and Todaro, M. P. 1970. Migration, unemployment and development: a two-sector analysis. *The American economic review*, 126–142.

Hijmans, R. J. 2019. *Geosphere: Spherical Trigonometry*.

Kapur, D.; and McHale, J. 2005. *Give us your best and brightest: The global hunt for talent and its impact on the developing world*. Center for Global Development Washington, DC.

Karney, C. F. F. 2013. Algorithms for Geodesics. *Journal of Geodesy*, 87(1): 43–55.

Kashyap, R.; and Verkroost, F. C. 2021. Analysing global professional gender gaps using LinkedIn advertising data. *EPJ Data Science*, 10(1): 39.

Laczko, F.; Tjaden, J.; and Auer, D. 2017. Measuring global migration potential, 2010–2015. *Global Migration Data Briefing Series*.

Lee, E. S. 1966. A theory of migration. *Demography*, 3(1): 47–57.

Lucchini, L.; Tonelli, S.; and Lepri, B. 2019. Following the footsteps of giants: modeling the mobility of historically notable individuals using Wikipedia. *EPJ Data Science*, 8(1): 36.

Massey, D. S.; Arango, J.; Hugo, G.; Kouaouci, A.; Pellegrino, A.; and Taylor, J. E. 1993. Theories of international migration: A review and appraisal. *Population and development review*, 431–466.

Melitz, J.; and Toubal, F. 2012. Native Language, Spoken Language, Translation and Trade. Working Papers 2012-17, CEPII.

Migali, S.; and Scipioni, M. 2018. A global analysis of intentions to migrate. *European Commission*.

Piore, M. J. 1979. *Birds of passage: Migrant labor and industrial societies*. Cambridge University Press.

Ravenstein, E. G. 1889. The laws of migration. *Journal of the royal statistical society*, 52(2): 241–305.

Raymer, J.; Wiśniowski, A.; Forster, J. J.; Smith, P. W.; and Bijak, J. 2013. Integrated modeling of European migration. *Journal of the American Statistical Association*, 108(503): 801–819.

Simini, F.; Barlacchi, G.; Luca, M.; and Pappalardo, L. 2021. A Deep Gravity model for mobility flows generation. *Nature communications*, 12(1): 1–13.

Skeldon, R. 2006. Interlinkages between internal and international migration and development in the Asian region. *Population, space and place*, 12(1): 15–30.

Spyratos, S.; Vespe, M.; Natale, F.; Weber, I.; Zagheni, E.; and Rango, M. 2019. Quantifying international human mobility patterns using Facebook Network data. *PloS one*, 14(10): e0224134.

Stark, O. 1978. *Economic-demographic interactions in agricultural development: The case of rural-to-urban migration*, volume 6. Food & Agriculture Org.

State, B.; Rodriguez, M.; Helbing, D.; and Zagheni, E. 2014. Migration of Professionals to the US. In *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11–13, 2014. Proceedings, Lecture Notes in Computer Science*, 531–43. Springer Cham, Switz.

The World Bank. 2020a. World Development Indicators. https://data.worldbank.org/indicator/SP.POP.TOTL. Accessed: 2021-12-01.

The World Bank. 2020b. World Development Indicators. https://data.worldbank.org/indicator/IT.NET.USER.ZS. Accessed: 2021-12-01.

Tjaden, J.; Auer, D.; and Laczko, F. 2019. Linking migration intentions with flows: Evidence and potential use. *International Migration*, 57(1): 36–57.

Willekens, F. 1994. Monitoring international migration flows in Europe. *European Journal of Population/Revue européenne de Démographie*, 10(1): 1–42.

Willekens, F.; Massey, D.; Raymer, J.; and Beauchemin, C. 2016. International Migration under the Microscope. *Science*, 352(6288): 897–899.

Willekens, F.; et al. 2019. Evidence-based monitoring of international migration flows in Europe. *Journal of Official Statistics*, 35(1): 231–277.

Wiśniowski, A. 2017. Combining Labour Force Survey data to estimate migration flows: The case of migration from Poland to the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1): 185–202.

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.; and Weber, I. 2012. You are where you e-mail: using e-mail data to estimate international migration rates. In *Proceedings of the 4th annual ACM web science conference*, 348–351.

Zagheni, E.; Weber, I.; and Gummadi, K. 2017. Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 721–734.

Zelinsky, W. 1971. The hypothesis of the mobility transition. *Geographical review*, 219–249.