Characterizing High-Skilled Mobility Patterns in Europe via LinkedIn

Keywords: high-skilled migration, social media data, LinkedIn data, statistical modeling

Extended Abstract

International high-skilled migration represents an increasingly large component of global migration streams with a significant impact on the global economy, gender imbalance in employment, and migration policies. Understanding the factors that explain why highly skilled workers move, and where they go, is of paramount importance in migration research, but generally difficult to measure and model. Timely, accurate, and comparative data on international migration are therefore of paramount importance for socio-demographic research and policy implementation, but generally not available or easily accessible via traditional data collections. In this context, the recent availability of large geo-located datasets has rapidly fostered research on high-skilled migration, for example, leveraging Facebook advertising data [4], LinkedIn advertising and recruiting platforms [2, 3], or bibliometric databases [5]. Properly harnessing such digital data sources would help grasp different nuances of this phenomenon and fill some of the gaps left by the traditional data.

In this study, we leverage a novel dataset collected via the LinkedIn advertising platform to characterize the mobility patterns of high-skilled workers in Europe. Specifically, LinkedIn Ads allows us to obtain an estimate of the number of LinkedIn users belonging to specific subgroups based on their characteristics, such as gender, age, or skills. Here we leverage this feature to retrieve the number of LinkedIn users that i) are highly skilled (i.e. having at least a BA, MA or PhD), and ii) moved from the country where they studied to another where they are currently employed. From this, we reconstructed the origin-destination matrix of the mobility flows of high-skilled workers between European countries, as shown in Figure 1A. Note that, for simplicity of language, here we refer to the number of LinkedIn users who moved from country i to country j as "flows" w_{ij}. In particular, 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 World Bank). This way we correct for potential biases due to under- or over-sampling the population, although population samples are already in good agreement (Spearman's $\rho = 0.91$, $p < 0.001$).

As shown in Figure 1A, 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 high-skilled migration. This is more evident in Figure 1B that shows the relationship between the inflows $w^{in} = \sum_{i \neq j} w_{ij}$ and the outflows $w^{out} = \sum_{j \neq i} w_{ij}$ by country. Specifically, for a given country i, if $w^{in} > w^{out}$ it means that, overall, the number of LinkedIn users that are currently employed in that country but studied elsewhere is greater than those who studied there and then moved elsewhere for their job, e.g., Luxembourg (LU), Switzerland (CH), Malta (MT), and the Netherlands (NL). On the other hand, when $w^{in} < w^{out}$ it means that the country lost more high-skilled migrants than those gained, e.g., Russia (RU), Poland (PL), and Romania (RO). Other countries, such as Spain (ES), Portugal (PT), Greece (GR), and Czechia (CZ), have instead roughly the same inflows and outflows.

In the absence of ground truth data that would be directly comparable with our definition of high-skilled migration, here we compare the total inflows w^{in} with the indicator given by the total employed foreign-born population from the International Labour Migration Statistics (ILMS). This comparison reveals that the mobility patterns estimated from LinkedIn are able to capture the observed ones (Spearman's $\rho = 0.93$, $p < 0.001$). However, many countries are missing in the ILMS dataset, thus making the LinkedIn estimates more comprehensive to characterize high-skilled migration across European countries.

After the descriptive analysis showing the relative attractiveness of high-skilled migration across countries, we employed a gravity-type model to identify those factors that drive such relationship [1]. Starting from the standard gravity model that assumes that movements depends on the populations at origin and destination and scale down with their distance, we can build up on this and add more variables to explore which ones actually play a role in shaping these movements. For this, we explored a variety of variables and employed a stepwise algorithm search approach with Bayesian information criterion to obtain the best model. Specifically, the final best model estimates that the number of users N_{ij} moving from origin country i to destination country j depend on the Linkedin population sizes at both origin and destination, scales down with their distance, are positively associated with language similarity between origin and destination, and with the indicator of GDP per capita and employment rates at destination. This means that high-skilled migration across countries in Europe is mainly driven by geographic and language proximity as well as favourable economic conditions at destination.

To conclude, here we presented the potential of using LinkedIn Ads data to characterize international migrations of highly skilled workers in Europe. Compared to traditional survey data, LinkedIn data have a number of advantages: they are continuously available, defined consistently across countries and languages, and provide a global and recent snapshot of high-skilled migration. On the other hand, social media data also comes with some limitations and challenges that we need to take into account to avoid limiting the validity of the conclusions drawn. Nevertheless, continuing to explore and propose alternative solutions is key to complementing traditional research methods and there's great potential to translate this frameworks into solid monitoring routines relevant to policy makers.

References

[1] 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.

[2] Perrotta, D.; Johnson, S. C.; Theile, T.; Grow, A.; de Valk, H.; and Zagheni, E. 2022. Openness to migrate internationally for a job: Evidence from LinkedIn data in Europe. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 759–769.

[3] Rodriguez, M.; Helbing, D.; Zagheni, E.; et al. 2014. Migration of professionals to the us. In *International conference on social informatics*, 531–543. Springer.

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

[5] Zhao, X.; Aref, S.; Zagheni, E.; and Stecklov, G. 2022. Return migration of German-affiliated researchers: Analyzing departure and return by gender, cohort, and discipline using Scopus bibliometric data 1996–2020. *Scientometrics*, 1–23.

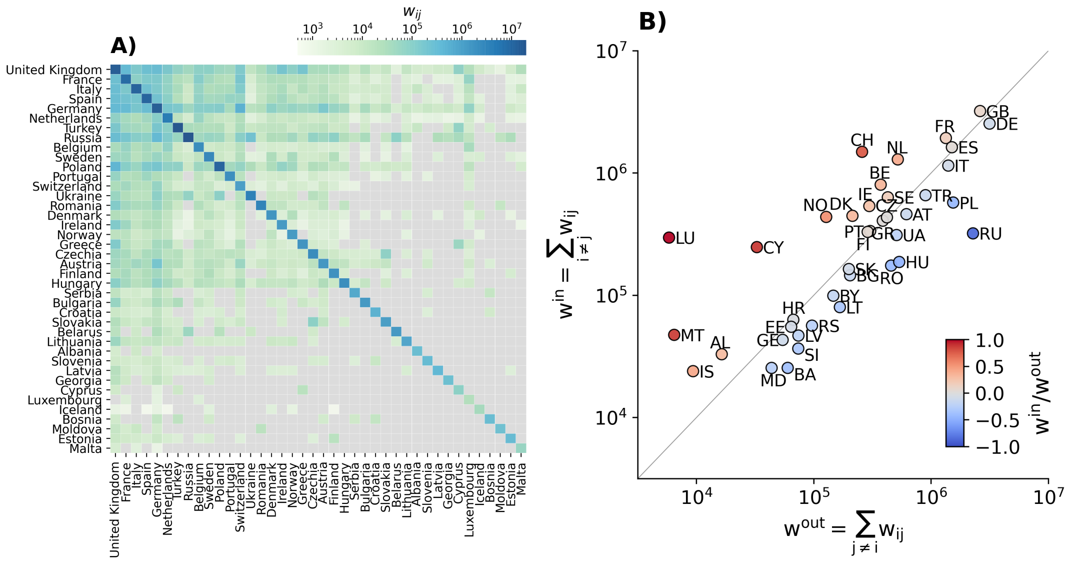

Figure 1: **High-skilled mobility patterns.** (A) Origin-destination matrix of high-skilled mobility flows w_{ij} of LinkedIn users who moved from country i where they studied (on the y-axis) to country j where they are currently employed (on the x-axis). The color code represents the normalized flows w_{ij} (grey indicates $w_{ij} = 0$). Countries are sorted according to LinkedIn population size. (B) Relationship between the inflows w^{in} and outflows w^{out} by country (self-loops excluded). The color 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.