Measuring Migration Conference
Nuffield College, University of Oxford | June 9-10, 2022
Session 1a: How do we measure migration? Methods and Advancements.

Use of non-traditional data sources to nowcast migration trends through Artificial Intelligence technologies.

Diletta Goglia[1]*, Laura Pollacci[1], Alina Sîrbu[1]

[1]Department of Computer Science, University of Pisa, Pisa, Italy
*Corresponding author: d.goglia@studenti.unipi.it

Keywords: International migration · migration flows · migration stocks · bilateral migration · Facebook social connectedness index · migration nowcasting · migration prediction · migration drivers

EXTENDED ABSTRACT

In recent years the pursuit of original drivers and methods is becoming an increasing requirement for migration studies, considering the new technologies used to characterise and understand the human migration phenomenon. Many researchers have proposed to employ non-traditional data sources to study migration trends, including so-called social Big Data such as online social networks [9, 1, 10, 12, 3, 8, 11, 2, 7, 5]. This unconventional approach is intended to find an alternative methodology to answer open questions about the human mobility framework (i.e., nowcasting flows and stocks, studying the integration of multiple sources and knowledge, and investigating migration drivers). The new data have the advantage of timeliness and large geographical coverage, but also disadvantages in terms of selection bias and the amount of resources required to process [6, 9]. Therefore, models extracted from these data need to be carefully validated, typically with traditional data sources. In this context of meaningful data combination, many types of data exist, still very scattered and heterogeneous, making integration far from straightforward.

Our work focuses on the integrated use of heterogeneous traditional datasets and new data types. We present two different contributions: use of alternative types of data and proposes two different models, a new multi-feature dataset [4] and a new indicator that could significantly contribute to the study of migration drivers and to forecast emerging trends through the use of Artificial Intelligence technologies.

We present a dataset to be exploited in migration studies as a concrete example of integration of traditional and non-traditional data sources: the Multi-aspect Integrated Migration Indicators (MIMI) dataset. It includes official data about bidirectional human migration (traditional country-to-country flow and stock data), multidisciplinary variables and original indicators, including economic, demographic, cultural and geographic indicators, together with the Facebook Social Connectedness Index (SCI). The integration process uniformised the data coming from various sources, in an attempt to fill in gaps and missing data, standardise location and time dimensions, and ultimately facilitate use by the research community. Thanks to this variety of knowledge, experts from several research fields (demographers, sociologists, economists) could exploit MIMI to investigate the trends in the various indicators, and the relationship among them. Moreover, it could be possible to develop complex models based on this dataset to assess human migration by evaluating related interdisciplinary drivers, and to nowcast/predict traditional migration indicators through non-traditional variables, such as the strength of social connectivity. Here, the SCI could have an important role. It guarantees an anonymised collection of information on users and their friendships, measuring the relative probability that two individuals across two countries are friends on Facebook. Therefore it could be employed as a proxy of social connections across borders to be studied as a possible driver of migration.

The second contribution is an analysis of the relation between indicators included in the MIMI dataset. We present a new measure, the Bidirectional Migration Probability (BMP) indicator, which takes into account both the inflows and outflows shared by two countries, and measures the relative probability of a person being a migrant from country *i* to *j* and vice versa. We predict the values of the BMP starting from SCI and other indicators, employing two different models. The first model is an ordinary least squares statistical model (OLS) that performs a linear regression to nowcast migration trends. Specifically, the model fits a subset of variables, including the Facebook SCI, and evaluates their relevance for nowcasting the BMP index. The migration drivers resulting from the fit of the OLS model will then be integrated with knowledge about past migration flows to build a second model, which consists of a Machine Learning Artificial Neural Network able to forecast migration.

All in all, our contribution lie in the need for new perspectives, methods, and analyses that can no longer prescind from taking into account a variety of new factors. The heterogeneous and multidimensional sets of data released with MIMI and exploited in the two models with the aid of the BMP indicator offer a new overview of the characteristics of human migration, enabling a better understanding and potential exploration of the relationship between migration and its drivers also through non-traditional sources of data.

Acknowlegdements

This work is supported by the European Union – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics", and by the Horizon2020 European projects "HumMingBird – Enhanced migration measures from a multidimensional perspective", Grant Agreement n. 870661.

References

[1] B. Claudio, G.-B. Sara, I. Stefano, M. Umberto, S. Francesco, and S. Spyridon. Data innovation in demography, migration and human mobility. 2022. DOI: 10.2760/958409.

[2] L. Fiorio, G. Abel, J. Cai, E. Zagheni, I. Weber, and G. Vinué. 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*, WebSci '17, page 103–110, New York, NY, USA, 2017. Association for Computing Machinery. DOI: 10.1145/3091478.3091496.

[3] C. Gendronneau, D. Yıldız, Y. Hsiao, M. Stepanek, G. Abel, S. Hoorens, A. Wiśniowski, E. Zagheni, L. Fiorio, and I. Weber. *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*. Publications Office of the European Union, 12 2019.

[4] D. Goglia. Multi-aspect Integrated Migration Indicators (MIMI) dataset, March 2022. DOI: 10.5281/zenodo.6493325.

[5] J. Kim, A. Sîrbu, G. Rossetti, F. Giannotti, and H. Rapoport. Home and destination attachment: study of cultural integration on twitter, 2021. DOI: 10.48550/ARXIV.2102.11398.

[6] L. Pollacci, L. Milli, T. Bircan, and G. Rossetti. Academic mobility from a big data perspective. 2022. PREPRINT (Version 1) available at 10.21203/rs.3.rs-1510153/v1.

[7] A. A. Salah. *Chapter 8: Mobile data challenges for human mobility analysis and humanitarian response*. Edward Elgar Publishing, Cheltenham, UK, 2021. DOI: 10.4337/9781839100611.00017.

[8] S. Spyratos, M. Vespe, F. Natale, I. Weber, E. Zagheni, and M. Rango. *Migration data using social media: a European perspective*. Publications Office of the European Union, 2018.

[9] A. Sîrbu, 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. Human migration: the big data perspective. *International Journal of Data Science and Analytics*, 11(4):341–360, 2021. DOI: 10.1007/s41060-020-00213-5.

[10] E. Zagheni, V. R. K. Garimella, I. Weber, and B. State. Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, page 439–444, New York, NY, USA, 2014. Association for Computing Machinery. DOI: 10.1145/2567948.2576930.

[11] E. Zagheni, K. Polimis, M. Alexander, I. Weber, and F. C. Billari. Combining social media data and traditional surveys to estimate and predict migration stocks. EAPS, 6 2018.

[12] E. Zagheni, I. Weber, and K. Gummadi. Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 43(4):721–734, 2017. DOI: 10.2307/26622775.