Comparison and Analysis of Migrant Identification Methods

L. Hong[1], *J. Wu*[1], *E. Frias-Martinez*[2], *A. Villareal*[3], *V. Frias-Martinez*[1]
(1) iSchool, University of Maryland, College Park, MD, USA, lzhong@umd.edu ,(2) Telefonica Research, Madrid, Spain. (3) Department of Sociology, University of Maryland, College Park, MD, USA.

Internal migration refers to the migration of individuals from one region to another within the same geopolitical entity, typically within the same country. Considerable attention has been given to the study of migration using spatio-temporal data passively generated, for example cell phone records or social media data. Such type of data enables to carry out large-scale analyses of migration flows as well as the micro-level view necessary to analyze individual behaviors [1]. At its core, research on migration behaviors first requires to identify the internal migrants in the dataset. Methods to identify migrants are based on determining home location changes i.e., a person that was living in a location, changes her home permanently or temporarily within the same country. Several methods have been developed to identify home location using spatio-temporal data [2] and some of these methods have been applied to identify volumes of internal migrants [3]. However, no work has looked into analyzing the impact that the choice of a home location algorithm might has in the identification of migrants. In this paper, we use an 8-month window cell phone dataset from Mexico to identify and compare internal migrations.

We define as internal migrants the individuals who have a consistent home location for at least three months and then move to another place, where they also stay for at least three months. With this definition, the internal migrants we identify can be either long-term or short-term (circular) migrants, depending on whether they go back or not to their original location after our data collection period finishes [4]. The census data we use for validation measures the internal migration flow at the municipality level. As a result, we use four different state-of-the art methods to detect home location at a municipality level: (1) most visited municipality at night, between 6pm to 6am; (2) a shorter period of night time, from 10pm to 6am- the main motivation to do this is that a tighter temporal range might help to reduce the noise in the set of communication towers considered as potential home location; (3) home location as the municipality where the center of gravity across all visited cellular towers is located, weighted by the cell phone activity in each cell and (4) a combination of the temporal window approach (10pm-6am) with the center of gravity approach- the assumption for this approach is that the activities during night hours tend to be closer to one's home location.

We used these methods to identify potential internal migrants and evaluated the accuracy of each method by comparing the computed migration matrix with official census data. Additionally, we looked into the biases introduced by each method, measured as the differences in accuracy when different types of urban and rural flows are approximated via cell phone data. Overall, results show that: (i) methods that use temporal ranges to identify internal migrants perform better than center of gravity-based methods; (ii) longer temporal ranges show better accuracy than shorter ranges; (iv) current methods show biases against rural population; and that (v) those biases decrease when total outbound or inbound flows are considered. Although we focused our study in Mexico, these results could potentially be valid to the great majority of developing economies, as they tend to show similar characteristics, namely: (1) high penetration of cell phones; (2) uneven distribution of the population between urban and rural areas; and (3) reduced cell phone activity when compared to developed economies.

References

[1] J. E. Blumenstock, "Inferring patterns of internal migration from mobile phone call records: evidence from rwanda," Information Technology for Development, vol. 18, no. 2, pp. 107–125, 2012.
[2] S. Isaacman, R. Becker, R. Caceres, S. Kobourov, J. Martonosi, M.and Rowland, and A. Varshavsky, "Identifying important places in peoples lives from cellular network data," in Int. Conf. Pervasive Computing, 2011, pp. 133–151
[3] S. Isaacman, V. Frias-Martinez, and E. Frias-Martinez, "Modeling human migration patterns during drought conditions in la guajira, colombia," in ACM Computing and Sustainable Societies, 2018.
[4] R. D. Bedford et al., New Hebridean Mobility: a study of circular migration. Canberra, ACT: Dept. of Human Geography, Research School of Pacific Studies, The Australian National University., 2017