Using IP Geolocation to Study Human Mobility

Bogdan State, Ingmar Weber, Emilio Zagheni
bstate@stanford.edu, ingmarweber@acm.org, emilio.zagheni@qc.cuny.edu

Motivation
Increasingly available geolocation data resulting from repeated logins to the same website offer the possibility of observing, for the first time, patterns of international mobility. Geolocation data have opened up new opportunities for theoretical advances in the field of migration studies, an area of research that has been strongly restricted by lack of data.

Data
Our paper uses data on the geographic locations from where over one hundred million anonymized users log into Yahoo! services to generate the first global map of short- and long-term mobility flows. We develop a protocol to identify a sample of hundreds of thousands anonymized users who spent more than three months in a different country from their stated country of residence, over a one-year period (“migrants”) and millions of users who spent less than a month in a different country from their identified country of residence (“tourists”). We compute aggregate estimates of migration propensities between countries, as inferred from a user’s location over the observed period.

Pendularity
Geolocation data allow us to characterize the pendularity of migration flows – i.e., the extent to which migrants travel back and forth between their country of origin and destination.

Determinants
We use data regarding visa regimes, colonial ties, geographic location and economic development to predict migration and tourism flows.

Conclusions
- We documented the persistence of global migration patterns explained by geographic, historical and economic factors.
- We observed the emergence of regional hubs of migration, like India and Brazil, alongside the US, which remains the global center in the network of migration flows.
- We noted that countries that experienced turmoil during the recent Arab Spring tended to have low levels of pendularity.
- The dataset, methods and results presented have important implications both for tourism advertisement and for several disciplines in the social sciences.

Acknowledgments
We are grateful to Yahoo! Inc. for providing us with access to their data and computing facilities. We thank Marko Radeta for his advice on map design. Bogdan State’s research was supported by the Joan Butler Ford Stanford Graduate Fellowship, and by the Amherst College Kellogg Fellowship.