Inferring International and Internal Migration Patterns from Twitter Data*

Emilio Zagheni
Queens College - CUNY & Wittgenstein Centre (IIASA, VID/ÖAW, WU)
New York City, NY, USA
emilio.zagheni@qc.cuny.edu

Venkata Rama Kiran Garimella & Ingmar Weber
Qatar Computing Research Institute
Doha, Qatar
vgarimella,iweber@qf.org.qa

Bogdan State
Stanford University
Stanford, CA, USA
bstate@stanford.edu

ABSTRACT

Data about migration flows are largely inconsistent across countries, typically outdated, and often inexistent. Despite the importance of migration as a driver of demographic change, there is limited availability of migration statistics. Generally, researchers rely on census data to indirectly estimate flows. However, little can be inferred for specific years between censuses and for recent trends. The increasing availability of geolocated data from online sources has opened up new opportunities to track recent trends in migration patterns and to improve our understanding of the relationships between internal and international migration. In this paper, we use geolocated data for about 500,000 users of the social network website "Twitter". The data are for users in OECD countries during the period May 2011- April 2013. We evaluated, for the subsample of users who have posted geolocated tweets regularly, the geographic movements within and between countries for independent periods of four months, respectively. Since Twitter users are not representative of the OECD population, we cannot infer migration rates at a single point in time. However, we proposed a difference-in-differences approach to reduce selection bias when we infer trends in out-migration rates for single countries. Our results indicate that our approach is relevant to address two longstanding questions in the migration literature. First, our methods can be used to predict turning points in migration trends, which are particularly relevant for migration forecasting. Second, geolocated Twitter data can substantially improve our understanding of the relationships between internal and international migration. Our analysis relies uniquely on publicly available data that could be potentially available in real time and that could be used to monitor migration trends. The Web Science community is well-positioned to address, in future work, a number of methodological and substantive questions that we discuss in this article.

*Emilio Zagheni worked on this article while he was a visiting researcher at the Wittgenstein Centre (IIASA, VID/ÖAW, WU), where he received helpful comments on this paper. In particular, we would like to thank Guy Abel and Nikola Sander.

WWW '14 *Companion*, April 7-11, 2014, Seoul, Korea

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Sociology; H.3.5 [**Online Information Services**]: Web-based services; H.2.8 [**Database Applications**]: Data mining

General Terms

Experimentation; Human Factors

Keywords

Migration; Twitter; Selection bias

1. INTRODUCTION

Migration is one of the major sources of demographic change [16]. Projecting migration rates and unveiling relationships between internal and international migration are thus important tasks for demographers and social scientists. Limited data availability has been one of the main bottlenecks for empirical analyses and for theoretical advances in the study of migrations. In particular, data about international migration flows are largely inconsistent across countries, typically outdated, and often inexistent [7, 8].

Demographers and international organizations are very interested in improving migration statistics. For instance, the European Statistical Agency EUROSTAT has worked in partnership with researchers at the University of Southampton and the Netherlands Interdisciplinary Demographic Institute to generate consistent estimates of migration flows between European countries [10].

One of the limitations for estimating and projecting migration flows is data availability for the recent past. Typically, there is a substantial lag between data collection and production of migration statistics. Even in those fortunate cases where it is possible to resolve inconsistencies between sources from different countries, it may take a few years before data become available, especially when the main source is a census.

The lack of timely data may strongly affect migration projections. Forecasts are particularly sensitive to recent trends. Therefore, extrapolations that fail to include information about the recent past may lead to much larger errors in the medium- and long-term.

The increasing availability of geolocated data from online sources has opened up new opportunities to identify migrants and to follow them, in an anonymous way, over time. In this paper, we use geolocated data from the social network website "Twitter" to evaluate recent trends in migrations in OECD[1] countries. The main goals

[1] www.oecd.org/about/membersandpartners/list-oecd-member-countries.htm

of our work are to complement existing migration statistics, and to develop methods for harnessing publicly available online data in order to improve migration forecasts and our understanding of populations of migrants.

The use of geolocated data for the analysis of human mobility has become more and more important during the last few years. Cellphone data have been used to analyze internal migration and mobility [5]. IP addresses from repeated logins to a website have been used to evaluate trends in international migration [24, 23]. Cellphone data often provide very detailed geographic information for a given country. IP addresses offer less accurate estimates of geographic location, but typically their scope extends beyond the borders of a single country. Geolocated Twitter data combine the best of two worlds. The geographic information (latitude and longitude) of Twitter data is very accurate and users post 'tweets' from different countries, before and after a potential relocation.

Twitter data allow us to look at migrations with a perspective that combines both internal and international mobility within one framework. Bridging the divide between internal and international migration is central to geographers and demographers. This article proposes an innovative way to address a major research question in social sciences with new online data. A second important aspect of this article is that the issue of selection bias when making inference from online data sources is addressed explicitly. Although we are not providing definitive answers, with this article we offer fresh perspectives and set the foundations for future interdisciplinary work.

In the next sections, after reviewing the most relevant related work, we explain how we created the dataset that we used to analyze migration and mobility in OECD countries. We provide a demographic description of the users in the sample. Then we discuss our methods and results about estimating recent trends and the relationships between internal and international mobility.

2. RELATED WORK

The study of human migration and mobility is not confined to a single field. Several lines of literature, that often cross disciplinary borders, have emerged. The increasing availability of geolocated digital records has led to a growing trend of interaction and exchange between scholars with different backgrounds.

Estimating flows of migrants and forecasting future trends is an important question, both to understand migration processes and for policy interventions. Abel [1] has developed statistical techniques to estimate flows of migrants from census data, in order to generate historical time series of migration flows. Among others, Abel's work is intended to inform the population projections of the International Institute for Applied Systems Analysis (IIASA). The Population Division of the United Nations (UN) has recently moved towards offering probabilistic population projections [20]. Forecasting migrations remains one of the most difficult tasks for the UN. Currently, there is a continuing collaboration between the UN and the University of Washington to develop statistical models to forecast net migration rates for all countries [2].

Statistical approaches to the study of human migration and mobility have been integrated within the framework of models used in physics. For instance, the most widely-known model for flows is a 'gravity' model [26, 7]. More recently, the 'radiation' model, an approach that addresses some of the limitations of gravity-type models, has been suggested [21].

The increasing availability of geolocated data from online sources has opened new opportunities to identify migrants and to follow them, in an anonymous way, over time [14]. Various types of data sources have been used to evaluate human mobility. Cellphone data have been used mainly to evaluate patterns and regularities of internal mobility for a country [3, 13, 6, 5], but also ties between countries (in terms of international calls) [4]. Travel itineraries for tourists have been inferred using geo-tagged pictures in Flickr [9] and recommendations posted on Couchsurfing [19]. Localized mobility, often within a city, has been measured using data from Twitter [12], Google Latitude [11], Foursquare [17] and public transport fare collection sensors [15, 22]. IP addresses have been used to evaluate internal mobility [18]. Analogously, recent trends in international flows of migrants have been estimated by tracking the locations, inferred from IP addresses, of users who repeatedly login into Yahoo! services [24, 23].

3. DATA AND METHODS

3.1 Data collection and pre-processing

For this project, we downloaded geolocated Twitter 'tweets' for about 500,000 users who have posted at least one geolocated tweet. The data set covers the period from May 2011 to April 2013.

We used the Twitter streaming API to search over posted tweets and considered only the subsample of geolocated tweets. We then mapped tweets (and the respective users) to countries, until we obtained geolocated tweets for about 3,000 users in each country considered. For this initial seed of users we downloaded all the geolocated tweets that we could obtain using the Twitter API. We then estimated the proportion of Twitter users whose geolocated tweets had been posted from exactly one single country, from two distinct countries, three countries, etc. The initial seed allowed us to set sample size goals for each country in order to obtain a balanced sample that accounts for different levels of geographic mobility. In other words, for countries where international geographic mobility is a relatively rare phenomenon, we oversampled. For countries with higher rates of mobility, we acquired relatively smaller samples. More specifically, for each country, we computed the fraction of users who had geolocated tweets in at least a different country from the 'home country'. We then sampled users with a probability inversely proportional to this fraction. For example, if in country A 50% of users post from more than one country, and in country B only 5% of users post from more than one country, then, for country B we would aim to obtain a sample of users that is 10 times larger than the one in country A. This would allow us to have a similar number of potentially mobile people in all countries considered, even in those countries where mobility is a rare phenomenon.

Starting with our initial seed, and following the sampling procedure described above, we obtained at least one geolocated tweet for a total of about 500,000 distinct users in OECD countries. For these individuals, we then downloaded all their public tweets. In this sample, about 345,000 had at least 10 geolocated tweets. About 150,000 posted at least 100 geolocated tweets. On average, users in the sample posted 142 geolocated tweets. The distribution is fairly skewed, with a median number of geolocated tweets equal to 34.

The average number of days between the first and the last geolocated tweet is 225. The average number of days between tweets is about 12. The average number of days between tweets reduces to about 6 for users who have posted at least 10 geolocated tweets.

There is a trade-off between a large sample of users for whom we may have sparse information over time, and a smaller sample of users for whom we have detailed and consistent information over time. We decided to select a sample of users for whom we have detailed and consistent geographic information since the early 2011. This decision is motivated by the fact that users for whom we have information over a longer period of time are more likely to provide reliable information. Moreover, they are more likely to continue to

post on Twitter and therefore we can follow them in the future. We split the dataset into separate periods of four months each, from May 2011 to April 2013, and we considered only those users for whom we have at least 3 geolocated tweets for each period of 4 months. The final sample size reduces to about 15,000 users.

3.2 Demographic characteristics of Twitter users in the sample

For each user we have, among others, a unique identifier, the text of their tweets, the date of the posts, and the geographic coordinates for the location from where the user 'tweeted'. Demographic information about users in the sample is not directly available. However, it can be estimated indirectly.

We used the face recognition software Face++[2] to estimate the gender and age of users based on their profile picture. Face++ uses computer vision and data mining techniques applied to a large database of celebrities to generate estimates of age and sex of individuals from their pictures.

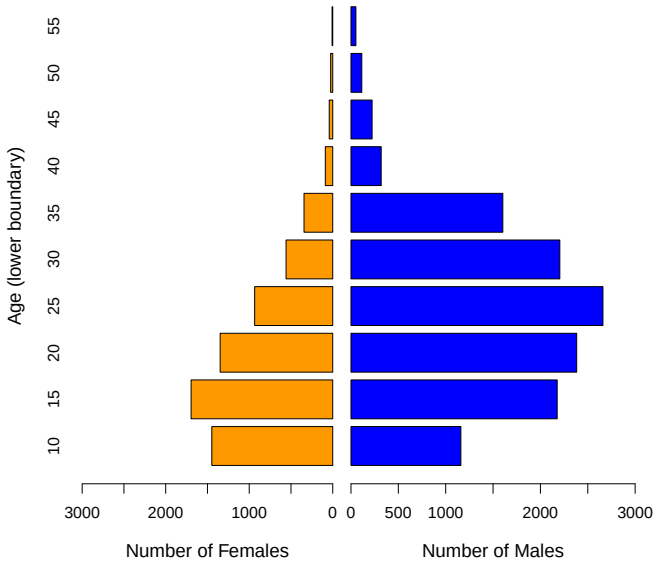

Figure 1: Population pyramid of Twitter users with a profile picture that could be evaluated.

Out of the sample of Twitter users with at least one geolocated tweet, we had 21,553 people with a profile picture that could be evaluated using Face++. Figure 1 shows a population pyramid for the users in the sample, based on Face++ estimates. About 63% of the users are classified as males. The average age for males is 26.2 years, whereas for females it is 19.7. The standard deviation is about 9.8 for both males and females.

The sample that we have is not representative of the whole population. However, the young age groups represented are particularly interesting, because migration and mobility are part of a broader set of transitions that happen mainly at young ages.

The estimates of age and sex that we presented are intended to provide a general idea of the population under study and should be considered with caution. There is quite a bit of uncertainty about estimates for single individuals. However, when the data is aggregated, as we did in the population pyramid, the uncertainty is substantially reduced, as overestimates and underestimates of age may cancel each other out. An important limitation of the approach is that estimates of age may be biased. For instance, users may have posted pictures of themselves when they were younger, or may not have updated their profile pictures. In this case we would generally underestimate the age of the population. Other text- and network-based methods to infer the age and gender could also be applied [25].

[2] http://en.faceplusplus.com

3.3 Estimation of trends in out-migration rates with a difference-in-differences approach

For each user for whom we have at least three geolocated tweets for each period of four months, we estimated the country of residence for the given period as the country from where most of the tweets were posted (the 'modal' country). If the uncertainty is large, i.e. if the number of tweets in the modal country is not at least three times as large as the number of tweets for the second most frequent country, then we discarded the information for that user for the specific period.

For a given user, if the modal country for two consecutive periods is the same, then we estimate that the user did not move over the period of eight months. If, for the first period of four months the modal country is A and then, for the second period, the modal country is B, we estimate that the user moved from country A to country B over the eight months considered.

The migration rates that we estimated cannot be considered representative of OECD countries. They represent the experience of migration and mobility of the subset of Twitter users who regularly post geolocated tweets. This population of Twitter users could be of great interest in itself. At the very least we are describing the experience of a fairly large and significant population. However, we would like to be able to use the information in our dataset of Twitter users to make some generalizations valid for the whole population.

We propose a difference-in-differences approach to estimate recent trends in mobility rates. Let m_c^t be the out-migration rate from country c to all other countries (number of users identified as migrants from country c/number of users in country c), at time t. Consider then the average of this quantity across all countries, m_{oecd}^t, i.e., the average of the migration rates at time t for all the OECD countries considered. If the population of Twitter users changes in similar ways across all the OECD countries over time, for instance due to Twitter's expanding user base, then we can use a difference-in-differences estimator to evaluate relative changes in trends:

$$\hat{\delta} = (m_c^t - m_{oecd}^t) - (m_c^{t-1} - m_{oecd}^{t-1}) \quad (1)$$

In other words, selection bias prevents us from making statistical inference for single points in time. In addition, changes in the composition of the Twitter population over time prevent us from using time series of estimated out-migration rates to make statistical inference about changes over time. However, if changes in the composition of Twitter users are consistent across countries, then the comparison of relative changes for a single country with relative changes for the group of reference can be used to provide information about trends. For example, if the proportion of Twitter users who are 25 years old is larger than the fraction of people who are 25 years old in the population, then estimates of migration rates based on Twitter data would tend to overestimate flows (because people in their 20s are more mobile than people in other age groups). Analogously, if the proportion of Twitter users who are 25 years old changes from one period to the next one, then we

cannot compare the two estimates from Twitter. However, if the proportion of Twitter users who are 25 years old changes in similar ways across all countries, then we can expect that for those countries where we observe more rapid increases in out-migration rates, the population-level migration rates have been increasing, relative to other countries.

3.4 Measuring the relationship between internal and international mobility

In addition to enriching predictions obtained through traditional demographic techniques, online datasets can be used to explore new questions related to mobility. Geolocated Twitter data include the geographic coordinates from where individuals post their tweets. The level of detail for the geographic information is very high and thus allows us to compute measures of mobility for users, and to evaluate patterns of mobility. In particular, we can distinguish between trajectories of mobility for those users that we classify as migrants and for those who continue to reside in the same country. This unique feature of the data set allows us to tackle one dimension of the important issue of the relationship between international and internal mobility and migration.

We use the radius of gyration as a measure of the distance covered by users over a certain period of time [5]. The radius of gyration is a measure of the average distance of geolocated tweets from their baricenter. In particular, we evaluate this quantity for subsets of the population (e.g., migrants vs. non-migrants in their home countries) to assess similarities and differences in the experiences across countries.

4. RESULTS

Figure 2 shows out-migration rates for selected countries, and the average of out-migration rates for OECD countries. These estimates were obtained using the methods described in the previous sections. The time series shown in Figure 2 are the starting point to generate difference-in-differences estimates of recent trends. The black line, which represents the average experience for OECD countries, cannot be taken at face value, because the sample is not representative of the whole population. However, country-specific relative departures from the general trend may be indicative of changes in mobility patterns.

Figure 3 shows estimates of difference-in-differences $\hat{\delta}$s for out-migration rates for OECD countries (for which we consistently have a sample of at least 100 Twitter users for each period of four months). The results shown are the average of $\hat{\delta}$s evaluated for the periods May-Aug and Sept-Dec of 2011, against the estimates for the respective periods in 2012. Positive values indicate a relative increase in out-migration rates.

The results provide interesting insights. For instance, they indicate a decline in out-migration rates from Mexico to other countries. Since the large majority of Mexican migrants move to the US, we can interpret the result as a sign of reduced migration and mobility from Mexico to the US. The result is consistent with recent estimates of the Pew Research Center for the period 2005-2010[3]. Twitter data show that the reversal in migration trends from Mexico, that occurred during the last few years, is persisting. This type of information would show up in official statistics only with a considerable delay. As a result, projections of migrations from Mexico to the US may tend to overestimate flows unless information about recent trends is incorporated.

[3] http://www.pewhispanic.org/files/2012/04/Mexican-migrants-report_final.pdf

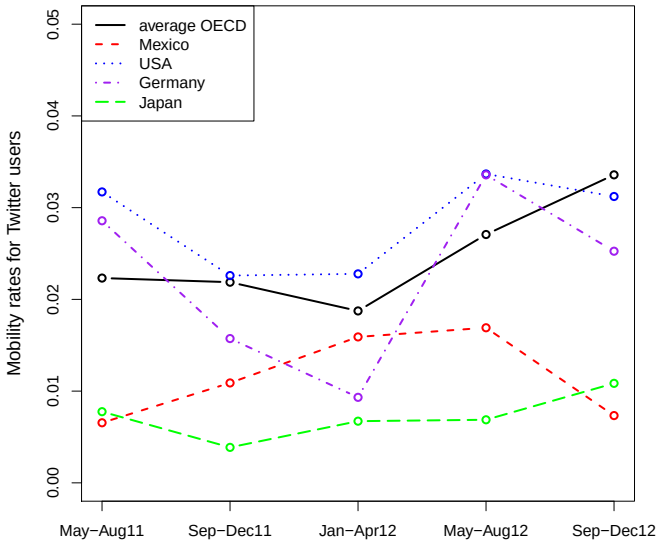

Figure 2: Out-migration rates for selected countries, and the average of out-migration rates for OECD countries. The rates are computed by evaluating the most frequent location of a user over the course of consecutive periods of four months.

Our results indicate that flows of migrants from the US have been slightly decreasing. This is in line with recent anecdotal evidence. In Southern Europe, Italy and Portugal seem to have reduced mobility towards other countries. For Spain and Greece we observe continued relative increases in out-migration rates. Analogously, migration from Ireland has been rising.

To evaluate the relationship between internal and international migration and mobility we estimated the radius of gyration of geolocated tweets (i.e., the average distance of tweets from their baricenter) for non-migrants and migrants, in their countries of origin. "Migrants" are those users that are identified as people who moved to a different country for at least one of the 4-month periods that we considered. Table 1 reports the results. We observed that the distance from the baricenter is larger for larger countries, as expected. However, for most countries, international migrants, when in their home countries, tend to travel shorter distances than people who did not migrate internationally. International migrants may tend to spend short periods of time in their home country, in their areas of origin. Those who do not migrate internationally may be more likely to move internally or to travel more in their home country. The US is a notable exception. The radius of gyration for international migrants in their home country is larger than the one for internal migrants or non-migrants. International migrants from the US are likely to be from a selected population of people who are highly mobile, both internationally and nationally. The result may also be related to a general reduction of internal geographic mobility in the US after the economic crisis.

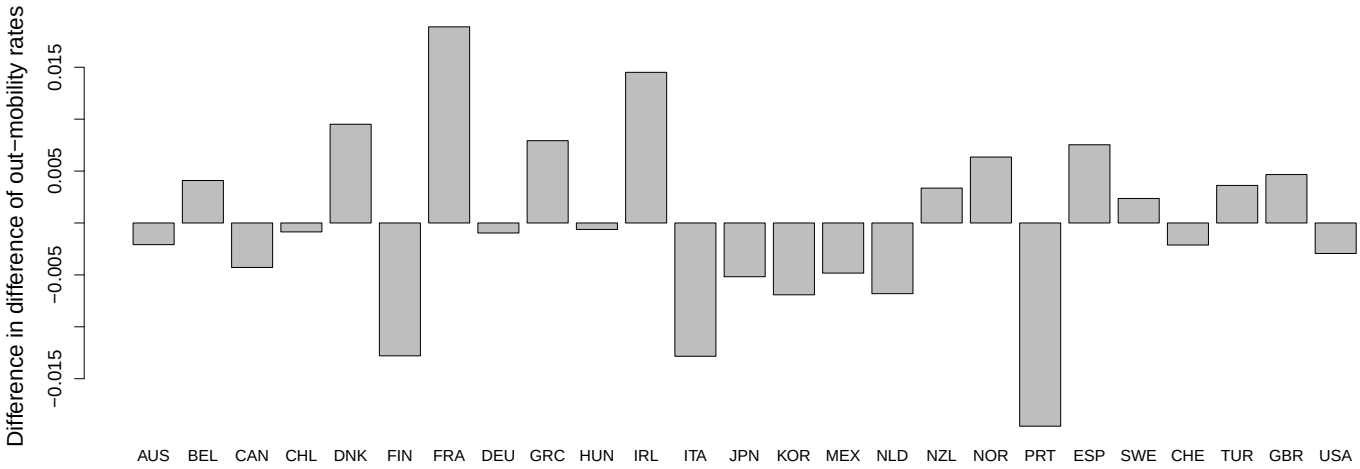

Figure 3: Values of the difference-in-differences estimator $\hat{\delta}$ for out-migration rates for OECD countries for which we consistently have a sample of at least 100 Twitter users for each period of four months.

Table 1: Radius of gyration in the home country for Twitter users identified as international migrants vs. those who were not classified as international migrants.

Country	Migrants	Non-migrants
Australia	7.664	9.719
Belgium	0.742	0.898
Canada	20.121	24.050
Chile	2.368	3.152
Denmark	1.457	1.761
Finland	2.249	2.031
France	3.141	4.556
Germany	3.091	3.677
Greece	1.754	1.863
Hungary	0.591	1.035
Ireland	1.724	1.517
Italy	2.781	3.435
Japan	2.598	3.721
South Korea	0.649	0.975
Mexico	3.535	5.470
Netherlands	1.069	1.030
New Zealand	1.698	3.900
Norway	1.838	3.240
Portugal	1.521	3.425
Spain	4.504	5.224
Sweden	3.288	3.847
Switzerland	1.328	1.397
Turkey	3.119	3.776
UK	1.970	2.316
US	27.204	23.007

5. CONCLUSIONS

In this article, we analyzed trends in mobility and migration flows using geolocated Twitter data. We provided a picture of very recent trends in OECD countries. We observed, among others, continued reductions of out-migration flows from countries like Mexico. In Europe, for some countries hit hard by the economic crisis, like Greece and Ireland, out-migration flows continue to increase. For others, like Italy, there are signs that out-mobility rates are decreasing, at least in relative terms. These observations are for short-term trends: there may be some stochasticity in the results and therefore the estimates should be taken with caution. Nonetheless, this type of information becomes available well before official statistics and thus can be seen as a "barometer" of mobility patterns that is useful to "nowcast" recent trends.

If we had data about very recent mobility patterns from official statistics, we could use that information to calibrate and validate our results. Those data are currently not available. As we collect more geolocated data from Twitter and generate longer time series of mobility rates, we expect to be able to calibrate our results against data from vital registration systems and migration surveys.

Our contribution is mainly methodological. In particular, we propose an approach to infer trends about mobility rates from biased samples obtained from a social media website like Twitter. Viewed from this perspective, our study can be considered as a feasibility study that opens up promising avenues of research for Web scientists and social scientists.

We proposed an approach that relies on difference-in-differences techniques to reduce the bias for statistical inference, we described the demography of our sample using a face recognition software, and we generated a new data set with observations that are not constrained by political borders. We believe that our work is a small, but important step towards addressing a central question for the Web Science community: how can we make statistical inference from online data when there is not any "ground truth" data that can be used as a training reference?

Going beyond what is well-known about human behavior is one

of the promises of computational social science. The Web offers data that contain relevant information about various aspects of social behavior. However, most social processes are very complex and hard to measure. Migration is a demographic process that is relatively easy to quantify. Still, the problem of operationalizing statistical inference for this well-defined process has not an obvious solution. Developing techniques to understand the population under study and to reduce the paramount problem of selection bias is an important task that goes beyond the substantive analysis of this article. We believe that understanding human behavior and social interaction via digital records of social media users requires the development of new tools. We expect the Web Science community to play a major role in addressing the issue of selection bias and in pushing the limit of what can be reliably inferred from online records.

6. REFERENCES

[1] G. J. Abel. Estimating global migration flow tables using place of birth data. *Demographic Research*, 28(18):505–546, 2013.

[2] J. J. Azose and A. E. Raftery. Bayesian probabilistic projection of international migration rates. *arXiv preprint arXiv:1310.7148*, 2013.

[3] M. A. Bayir, M. Demirbas, and N. Eagle. Discovering spatiotemporal mobility profiles of cellphone users. In *World of Wireless, Mobile and Multimedia Networks & Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*, pages 1–9. IEEE, 2009.

[4] J. E. Blumenstock. Using mobile phone data to measure the ties between nations. In *Proceedings of the 2011 iConference*, iConference '11, pages 195–202, New York, NY, USA, 2011. ACM.

[5] J. E. Blumenstock. Inferring patterns of internal migration from mobile phone call records: Evidence from rwanda. *Information Technology for Development*, 18(2):107–125, 2012.

[6] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.

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

[8] J. De Beer, J. Raymer, R. Van der Erf, and L. Van Wissen. Overcoming the problems of inconsistent international migration data: A new method applied to flows in europe. *European Journal of Population*, 26(4):459–481, 2010.

[9] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu. Automatic construction of travel itineraries using social breadcrumbs. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, pages 35–44. ACM, 2010.

[10] R. v. d. Erf. Analysis of final results. In *Paper prepared for the IMEM progress meeting on 22-24 February in Asker, Norway*, 2012.

[11] L. Ferrari and M. Mamei. Discovering daily routines from google latitude with topic models. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 432–437. IEEE, 2011.

[12] L. Ferrari, A. Rosi, M. Mamei, and F. Zambonelli. Extracting urban patterns from location-based social networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pages 9–16. ACM, 2011.

[13] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[14] P. Hui, R. Mortier, M. Piorkowski, T. Henderson, and J. Crowcroft. Planet-scale human mobility measurement. In *HotPlanet*, page 1, 2010.

[15] N. Lathia, D. Quercia, and J. Crowcroft. The hidden image of the city: sensing community well-being from urban mobility. In *Pervasive Computing*, pages 91–98. Springer, 2012.

[16] R. Lee. The outlook for population growth. *Science*, 333(6042):569–573, 2011.

[17] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An empirical study of geographic user activity patterns in foursquare. *ICWSM*, 11:70–573, 2011.

[18] A. Pitsillidis, Y. Xie, F. Yu, M. Abadi, G. M. Voelker, and S. Savage. How to tell an airport from a home: Techniques and applications. In *Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks*, page 13. ACM, 2010.

[19] E. Pultar and M. Raubal. A case for space: physical and virtual location requirements in the couchsurfing social network. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, pages 88–91. ACM, 2009.

[20] A. E. Raftery, N. Li, H. Ševčíková, P. Gerland, and G. K. Heilig. Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109(35):13915–13921, 2012.

[21] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

[22] C. Smith, D. Quercia, and L. Capra. Finger on the pulse: identifying deprivation using transit flow analysis. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 683–692. ACM, 2013.

[23] B. State, I. Weber, and E. Zagheni. Studying inter-national mobility through ip geolocation. In *WSDM*, pages 265–274, 2013.

[24] E. Zagheni and I. Weber. You are where you e-mail: using e-mail data to estimate international migration rates. In *WebSci*, pages 348–351, 2012.

[25] F. A. Zamal, W. Liu, and D. Ruths. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. In *ICWSM*, 2012.

[26] G. K. Zipf. The p1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946.