Characterization of Internal Migrant Behavior in the Immediate Post-Migration Period using Cell Phone Traces

Lingzi Hong
iSchool, UMD
lzhong@umd.edu

Jiahui Wu
iSchool, UMD
jeffwu@umd.edu

Enrique Frias-Martinez
Telefonica Research
e.friasmartinez@telefonica.com

Andrés Villarreal
Department of Sociology, UMD
avilla@umd.edu

Vanessa Frias-Martinez
iSchool, UMD
vfrias@umd.edu

ABSTRACT

Internal migrations have been studied using two types of approaches: macro-level and micro-level analyses. Macro-level studies are typically carried out using a combination of various survey and census datasets to model large-scale behaviors, however these models fail to provide more nuanced information about the physical or social status of the migrants. Micro approaches, which successfully use interviews and diaries to provide a window into more individual behaviors, could benefit from methods to identify novel or under-studied behaviors that should be addressed in the migration research agenda. In this paper, we present a framework that uses information extracted from cell phone metadata to reveal internal migration behaviors that could guide or complement the research agenda of micro-level migration researchers working to understand the physical, social and psychological decision processes behind migration experiences. The proposed framework allows to carry out micro-level analyses of internal migration with a focus on immediate post-migration behaviors and the role of pre-migration activities from two perspectives: spatial behaviors and social ties. Ultimately, we expect our analyses to inform migration researchers of pre- and post-migration behaviors that would benefit from further qualitative analysis.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**;

KEYWORDS

cell phone data, spatial dynamics, social relationships

ACM Reference Format:
Lingzi Hong, Jiahui Wu, Enrique Frias-Martinez, Andrés Villarreal, and Vanessa Frias-Martinez. 2019. Characterization of Internal Migrant Behavior in the Immediate Post-Migration Period using Cell Phone Traces. In *Tenth International Conference on Information and Communication Technologies and Development (ICTD '19), January 4–7, 2019, Ahmedabad, India.* ACM, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/3287098.3287119

ICTD '19, January 4–7, 2019, Ahmedabad, India

ACM ISBN 978-1-4503-6122-4/19/01…$15.00
https://doi.org/10.1145/3287098.3287119

1 INTRODUCTION

Internal migration refers to the migration of individuals from one region to another within the same geopolitical entity, typically within the same country [2, 32]. In recent years, there has been an increase in the volume, types and complexity of human internal migration in many countries, mostly due to economic crises, political instability and various types of natural disasters [54]. Economists, demographers, geographers and sociologists have extensively studied internal migration movements mostly focusing on two areas: causes (determinants) and consequences of migrations [25]. Understanding the causes and the consequences of internal migrations is critical for the development of policies that guarantee that migrants are not left behind and that allow to enhance their adaptation to the new settings so as to enhance their work opportunities, lifestyle and family relationships [31].

Determinants and consequences of internal migrations have been studied using a large variety of frameworks that can be grouped into two approaches: macro-level and micro-level analyses. Macro-level studies are typically carried out using a combination of various survey and census datasets, including origin-destination internal flows as well as demographic and socio-economic data, to assess the role that specific variables might play as both determinants or consequences of internal migration movements [7]. These macro-level studies provide general migration trends that are highly useful from a policy perspective. Decision makers can assess the types of social groups, characterized for example by age or profession, that are migrating between regions, and the long term impact that these migrations have, for example, on the local economy. However, the macro-level analyses fail to evaluate more nuanced variables that can be captured through micro-level analyses including data from interviews and diaries [55].

Micro-level analyses provide a window into the physical, social and psychological status of internal migrants showing, for example, that migrants maintain strong social ties with the communities they leave behind [27]; that migrants show different spatial behaviors to locals, mostly due to search processes in the physical environment that generate spatial dynamics that differ from those that locals show [9]; or that internal migrants usually encounter difficulties in adapting to new environments, suffering from a series of issues such as psychological stress which might affect their behavior in the physical environment [36, 42, 65]. Although the migration research agenda at the micro-level is broad, scholars recognize the need to better grasp the complexities of migration behaviors, its causes and determinants through new lines of inquiry [6]. In this paper, we present a framework that uses information extracted from cell phone metadata to reveal internal migration behaviors that could guide or complement the research agenda of micro-level migration researchers working to understand the physical, social and psychological decision processes behind migration experiences. The ubiquitous generation of data, such as the one produced by location-based social networks (LBSN), social media or cell phones, provides an opportunity to collect and analyze internal migrants' behaviors passively at a scale that was not possible before. In the particular case of cell phone metadata, given the large penetration rates of cell phone usage worldwide, it enables us to study the aggregated behavior of internal migrants at a country scale. Furthermore, since cell phone metadata contains the georeferenced locations visited by an individual and her contacts, it also allows to model individual behaviors both in terms of spatial dynamics as well as social networks.

The proposed framework allows to reveal internal migration behaviors with a focus on immediate post-migration behaviors and the role of pre-migration activities from two perspectives: spatial dynamics and social ties. The main objective is to carry out large-scale analyses of internal migration trends so as to reveal individual migrant behaviors that would benefit from further qualitative studies through personal interviews or individual surveys. Ultimately, we expect our analyses to inform migration researchers of pre- and post-migration behaviors that would benefit from further qualitative analysis. Given cell phone metadata from millions of individuals for a given country, the proposed framework consists of three parts. First, the framework uses features extracted from the cell phone metadata to identify potential migrants in the dataset. We present a method to identify internal migrants and we evaluate its accuracy using real census migration data. Second, the framework uses the identified migrants to characterize immediate post-migration behaviors *i.e.,* we analyze the spatial dynamics and social networks of migrants post-migration, and compare these against behaviors from locals that have not undergone any migration process. Third, we analyze the role that pre-migration spatial dynamics and social networks might play in the same post-migration behaviors shown by internal migrants. We evaluate the proposed framework to study internal migration behaviors in Mexico, using a dataset with eight months of anonymized cell phone metadata from over 48 million subscribers; and we show how our findings could complement future qualitative studies in Mexican internal migration.

2 RELATED WORK

Survey-based Migration Analysis. Survey-based studies on internal migration have focused on the use of macro- and micro-approaches to assess pre- and post-migration behaviors brought up by migratory movements [10, 18]. The features used in macro approaches typically differ from the micro approaches due to the nature of the datasets they use: while macro approaches typically focus on large-scale census data, micro approaches are prone to use survey and interview data at smaller population scales. Macro-approaches have shown, using census data, that internal migrants tend to relocate due to unemployment, lack of services, poverty or lack of safety to areas that offer better conditions [45, 58]. These features have been used by researchers, to build theoretical models that explain and predict migrations at the macro scale, including Zipf's inverse distance law [57], Stouffer's law of intervening distances [56] or Gravity models [49]. In this paper, we will use some of these theoretical models to evaluate the quality of the proposed method to identify migrants in a cell phone dataset.

On the other hand, micro approaches have mostly focused on analyzing behavioral features that characterize the mobility, sociality or psychology of migrants. The studies with behavioral features closest to ours (spatial dynamics and social networks) are numerous, including Li *et al.* which showed that migrants in Beijing have restrained spatial dynamics which are heavily skewed towards people from the same town of origin [41]; Nogle *et al.* that analyzed the importance of offering migrants settlement assistance as well as information about local opportunities [47]; Hendriks *et al.* observed that migrants tend to spend less time in social activities [29]; Gurak *et al.* who showed that migrants' life status can be affected by factors such as such as kin and friend relationship [27]; Kuo *et al.* who related the adaptation problems of migrants to the difficulties in the restoration of disrupted social networks [37]; or Chib *et al.* who showed that migrants rely on social support (emotional, instrumental or information aid) to deal with the stress caused by their migration experiences [12]. In this paper, we focus on the quantitative analysis of internal migration behaviors with a focus on spatial dynamics and social relationships. We explore automatic methods to capture behaviors via cell phone data; and we analyze how some of our results could guide or complement current qualitative internal migration studies.

Migration Analysis with Passively Collected Data. Ubiquitous data generated in a passive manner has been used in a much limited number papers, but with a focus on macro analyses that model the volume and direction of the migrations flows [63]. Zagheni *et al.* used email service logs to identify international migration rates [64]. Weber *et al.* used anonymized log data from Yahoo! services users to generate short-term and medium-term migration flows across countries [61]. Given the geographic accuracy of IP addresses, the authors focused on international migration rates. Similarly, geolocated Twitter data can reveal users' location at different scales: from a city to a GPS location. As a result, this type of data has been used to model both international and internal migration patterns at a macro level, both in terms of flow volumes and directions [63].

Nevertheless, all these approaches suffer from a large bias problem, since the demographic and economic backgrounds of email, web and twitter users is not representative of the population at large [59]. On the other hand, cell phone data, with much higher penetration rates across all types of people, has been shown to be more representative, although still imperfect, of the population at large [21]. As a result, Blumenstock *et al.* proposed a macro-level method that used cell phone metadata to identify migrants and quantify volumes and directionality of internal migrations in Rwanda [5]. As shown, all these studies with passively collected data focus on macro approaches that characterize the volume of migration flows. This paper extends the state of the art with a framework that allows to carry out individual internal migrant behavior analyses using cell phone metadata. The framework characterizes migrants' behaviors in terms of spatial dynamics and social relationships to analyze immediate post-migration behaviors and the role of pre-migration behaviors on post-migration activities.

3 RESEARCH QUESTIONS

We present a framework that uses cell phone metadata to quantitatively analyze internal migrant behavior in the immediate post-migration period and to evaluate the role that pre-migration behaviors might play on the immediate post-migration activities of internal migrants. Whenever applicable, we will also frame our quantitative findings within ongoing qualitative migration studies so as to assess the validity of our framework in identifying migrant behaviors of interest to the migration research community. We focus our analyses on two types of behavioral features: spatial dynamics and social ties. Spatial dynamics refer to the mobility patterns that individuals have. These variables can reveal insights with respect to the spatial dispersion or spatial diversity of migrants, throwing light into the impact that migrations might have in how individuals explore their new physical environment. On the other hand, social features will provide information about the social ties of migrants, before and after migration, which might help to analyze the role that social relationships might play in the immediate adaptation of migrants to their new environment. To analyze the spatial dynamics and social networks of internal migrants, we first need to identify the migrants themselves in a cell phone dataset. We will present a method that uses cell phone metadata to identify long-term changes in home locations, and evaluate its accuracy against real, census-based, internal migration data. The framework will contribute with quantitative methods to answer the following two research questions (see Figure 1 for details):

RQ1. Characterization of immediate post-migration behaviors. We present a method to analyze spatial dynamics and social networks in the immediate post-migration period and to compare these against local individuals so as to assess the behavioral differences potentially due to the migration. This analysis will reveal insights that might be helpful for future qualitative studies focused on understanding how internal migrants adapt to their new settings, and on the difficulties that they might find in that adaptation process.

RQ2. Analysis of the role that pre-migration behaviors have on post-migration activity. We describe a multivariate regression method to examine and quantify the role that the pre-migration spatial dynamics and social networks have in the post-migration behaviors that migrants show immediately after they arrive to their new locations. The model described can potentially be used as a tool to foresee the types of spatial and social behaviors that internal migrants might have, given a specific type of pre-migration population. Further qualitative research in this direction, could translate into the development of policies to better assist and ease the adaptation of migrants to their new surroundings [16]. We explore the applicability of the proposed framework and methods using cell phone metadata for the country of Mexico; and present behavioral findings that illuminate the spatial dynamics and social relationships that internal migrants have.

4 CELL PHONE DATA DESCRIPTION

To characterize migrant spatial dynamics and social ties, we use the information collected by a cell phone network. Cell phone networks are built using a set of base-transceiver stations (BTS) that are responsible for communicating cell phone devices within the network. Each BTS or cellular tower is identified by the latitude and longitude of its geographical location. The coverage area of each individual BTS is called a cell. The geographical area covered by a cell mainly depends on population density, and typically ranges from less than 1 km^2, in dense urban areas, to more than 4 km^2, in rural areas. For simplicity, it is common in the literature to assume that the cell of each BTS is a two-dimensional non-overlapping polygon, which is typically approximated using Voronoi diagrams [20, 50].

Call Detail Records (CDRs) are generated by telecommunication companies for billing purposes. CDRs are created whenever any type of cell phone connected to the network makes or receives a phone call or uses a service (e.g., SMS, MMS). In the process, the BTS details are logged, which gives an indication of the geographical position of the cell phone at the time of the call. Note that no information about the exact position of a user in a cell is known, *i.e.,* we do not have a GPS-type location of the phone within the coverage area of a BTS. CDR data can be used to model spatial dynamics and social ties. Spatial dynamics are characterized with features that use the geographical position of the cellular towers used for placing calls, while social ties can be characterized using the people that an individual talks to. Sections 6 and 7 will describe in depth the specific features we use in our analyses.

Here, we use an eight-month anonymized CDR dataset for the country of Mexico (October 2009 to May 2010). To preserve privacy, original records are encrypted and all the information presented in the paper is aggregated. From all the information contained in a CDR, our study considers the encrypted originating number, the encrypted destination number, the time and date of the interaction, and the BTS that the cell phone was connected to when the call was placed. No contract or demographic data was considered or available for this study and none of the authors of this paper collaborated in the extraction and the encryption of the original data. The dataset contains 7 billion records from 39K cellular towers that cover the whole country. We eliminate from the dataset all cell phones (and their corresponding CDR data) whose activity can be assumed to correspond to a machine and not an individual using the approach in [48]. This approach, which uses average measures of reciprocal cell phone contacts and frequency to eliminate anomalous accounts, was applied over the dataset leaving a final number of 48M unique cell phones in the dataset.

5 MIGRANT IDENTIFICATION

In this section, we present a method to automatically identify internal migrants in a CDR dataset; and propose two approaches to validate that the migrants and migration flows identified can be used as representatives of the country migration flows at large. We carry out the validation by comparing the internal migrants and their migration flows against actual census data and report their similarity.

Since we only have an eight-month observation window with CDR data, we define as internal migrants individuals that 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. A similar method was described by Blumenstock *et al.* in [5]. With this definition, the internal migrants we identify can be either long-term or short-term (circular) migrants depending on whether they

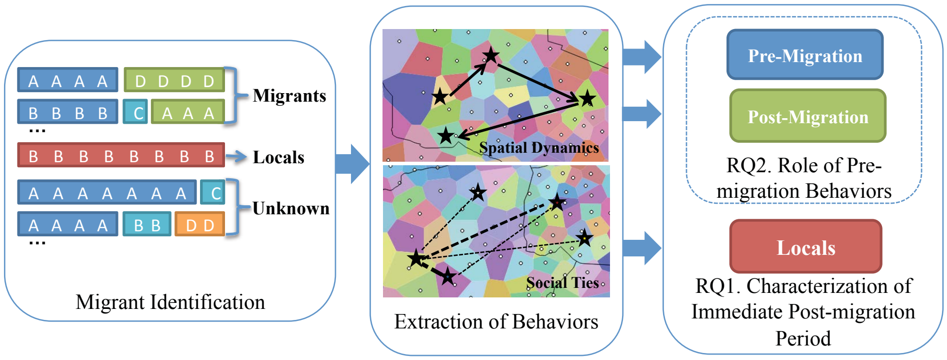

Figure 1: Overview of the proposed framework and research questions.

go back or not to their original location some time after our data collection period finishes [4]. A limitation of our work is that we will not differentiate between these two types of internal migration. However, since we compare this method against internal migration census data that does not differentiate between the two, our validation will still be consistent.

Home location algorithms are contained within a larger group of algorithms used to identify the important or meaningful places of an individual from their mobility information. Although the bulk of the state of the art focuses on CDR, the algorithms can be used for any set of non-continuous location traces. In the literature, important places are typically classified as home, work or other [46]. The main idea behind these algorithms consists in using some criteria to define time slots for home, work and other activities and then use the mobility information available to identify the location of these important places. The most well known approaches include Ahas *et al.*, who used an anchor-point model to identify home and work and validated it with the actual geography of the population finding a high level of correlation [1]; the work by Isaacman *at al.* who clustered cellular towers (or active points) to identify home and work [35]; or the work by Frias-Martinez *et al.* who proposed a genetic algorithm to identify the time slot that had to be used to better characterize home and work [22].

Based on the state of the art described, we identify the daily pre- and post-migration home location of an internal migrant as the most used BTS tower between the hours of 6pm and 6am Monday through Thursday each week, with the assumption that it is highly probable that a person will be home at night and in week days. We also explored other home location methods based on the center of gravity of the most used BTS towers as presented in [5, 35]. However, the migration validation that we discuss in this section showed that the best results were obtained for the time-range method. Finally, we assign home locations at the municipality level because the census data that we use to validate our method measures internal migration at that granularity level. To identify internal migrants and their flows, we need to determine the individuals whose home location was at a given municipality for at least three months, and then changed to another municipality and remained the same for at least another three months. To achieve that, we compute home location (municipality) as explained, for each individual on a daily basis. If there is no information to identify a home during a week day, the last position identified for week day was assigned. However, any user with at least one week without a home location assigned is not considered as potential migrant due to a lack of mobility information. Once all daily home locations have been computed, if at least 70% of the municipalities identified as home location are the same, the individual is assigned that home for the month. Finally, we search for individuals whose municipality home location throughout the eight months of data changes only once, with at least three months in each municipality. Following this procedure, we identify a total of 18,580 internal migrants in the dataset.

5.1 Validation

We conduct two validation experiments to verify the consistency between the migration flows we detect with our method, and the real migration flows.

Validation 1. We use the *real* migration flow matrix computed by the Mexican Statistical Institute (INEGI) which is based on the ENADID 2010 survey (National Survey of Demographic Dynamics) [34]. The survey records the number of people migrating from one municipality to another from 2005 to 2010 across all municipalities in Mexico. We use the internal migrants identified by our method to compute our own CDR-based migration flow matrix, and compare it against the official one via Pearson's correlation analysis between the two matrices. Specifically, we compute the correlations for: (i) the migration between each pair of municipalities *i.e.,* correlation between each pair (origin,destination) in the flow matrix, (ii) the outbound migration across municipalities *i.e.,* correlation between each pair (origin,all destinations) in the migration flow matrix and (iii) the inbound migration across municipalities *i.e.,* correlation between each pair (destination, all origins) in the migration flow matrix. The results show a strong correlation between the real and the CDR-based matrices with correlation coefficients of .60, .82 and .74, respectively (with all $p < 2.2e-16$).

Validation 2. The previous validation compared the official internal migration flows across five years against the CDR-based migration flows obtained with eight months of data from 2009-2010. As such, the correlation coefficients might be affected by the disparity in the data collection periods. To overcome this limitation, we present a second validation that, rather than comparing actual migration behaviors against CDR-based migration behaviors, compares the similarity between the outcomes of migration models

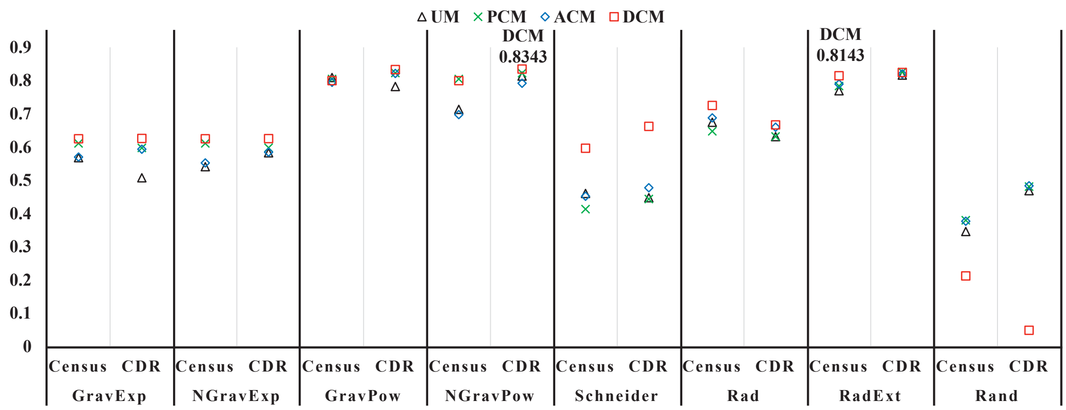

Figure 2: Goodness of fit with census-based and CDR-based Migration Flows. The migration models explored are GravExp, NGravExp, GravPow, NGravPow, Schneider, Rad and RadExt. The constraints are UM, PCM, ACM and DCM.

and the two datasets. By comparing similarities between model outcomes and actual migration volumes, we expect to minimize the impact of diverse temporal windows in the analysis. There exist two main families of migration models to explain internal or international migration patterns: Gravity models [67] and Schneider's Intervening opportunities models [56] (the latter have evolved into Radiation models [53]). While gravity models assume that the volume of migrations between two locations decreases with distance, intervening opportunities models assume that the migration flow between two locations depends on the number of opportunities that each location has *i.e.,* the decision to migrate is not related to the distance between the two places, but rather to the possibilities of settlement at the destination location. The two types of migration models have been used in the past to explain both migrations and commuting patterns with various levels of success [38, 39, 53].

We fit these two types of models using: (i) distances between origin and destination municipalities for the gravity models; and (ii) population of the municipalities [33] as as proxy for intervening opportunities (as is currently done in the state of the art for migration models [53]). Once these models are built, we compare its *theoretical* outcome against the census-based migration flows and the CDR-based migration flows and report prediction accuracies. We measure similarities between the models and the flows using the common part of commuters feature (CPC) [38] which measures the percentage of correct predictions for the number of people that migrate to other municipalities. It varies from 0, when no agreement is found, to 1, when the two migration flows are identical. Specifically, we fit the following migration models: Gravity law model with an exponential distance decay function (*GravExp*); Normalized gravity law model with an exponential distance decay function (*NGravExp*); Gravity law model with a power distance decay function (*GravPow*); Normalized gravity law model with a power distance decay function (*NGravPow*), Schneider's intervening opportunities law model (*Schneider*); Radiation law (*Rad*) and Extended radiation law models (*ExtRad*). Each one of this migration models considers four different constraints: Unconstrained model, or *UM*, that just preserves the total number of trips; Production constrained model, or *PCM*, that assumes that the number of trips produced by a geographical unit are preserved; Attraction constrained model, or *ACM*, that assumes that the number of trips attracted by a unit are preserved and Doubly Constrained Model, or *DCM*, that assumes that both the trips generated and attracted are preserved (see [38, 40] for further details).

Figure 2 shows the CPC results for both the census-based migration flows and the CDR-based migration flows for each migration model and each constraint. The extended radiation law model (RadExt) [62] is the one that performs the best with a relatively high goodness-of-fit of 0.8143 for the census-based migration flow; while for the CDR-based migration flow the normalized gravity law (NGravPow) performs best with a goodness-of-fit of 0.8343. These results indicate that the set of internal migrants identified using the method proposed in this section appears to represent well the real internal migration flows at the country scale. Next two sections will use these internal migrants to analyze immediate post-migration behaviors and the role of pre-migration behaviors from a spatial dynamics and social ties perspective.

6 CHARACTERIZATION OF IMMEDIATE POST-MIGRATION BEHAVIORS

In this section, we analyze the internal migrant behavior in the immediate post-migration period from two perspectives, spatial dynamics and social ties; and compare them against the behavior of the local population in their final destinations. Such comparison, will allow us to identify behavioral differences that could be potentially explained as consequences of the internal migration process, rather than as artifacts of the new physical environment *i.e.,* a migrant might change her commuting patterns post-migration but this could be due to the fact that one is migrating to a city with a different urban geography. In fact, related work has shown that spatial dynamics and social networks are often times *shaped* by their physical environment [13, 19, 24]. Thus, by looking at differences with local behaviors we expect to discern between changes due to the physical environment or issues that migrants face and that are specific to their migrant community. We identify the local population (locals) in the CDR dataset as individuals who have a home location assigned to a municipality where internal migrants migrate to, and whose home location is the same throughout the length of our dataset (eight months). Using the home location algorithm explained before and these requirements, we end up with $1,505,868$ local residents across all municipalities that we have identified as migrant destinations. To assess that the local population sampled as local can be used as a representative of the population at large, we compute Pearson's correlation between the real population numbers per municipality and the population sampled as local for that municipality. The two values are consistent across most municipalities, with a high Pearson's correlation coefficient of 0.8570, and a Spearman's coefficient of 0.8295 (with $p < 2.2e-16$ for both), showing a high rank correlation as well.

Our final objective is to analyze the immediate post-migration behaviors of migrants and compare these against the behaviors of locals to assess migrant behaviors potentially consequence of the migration process itself and not of the physical environment they migrate to. We carry out this comparison by modeling spatial dynamics and social network features for both migrants (pre- and post-migration) and locals with the CDR data available. Next, we identify the migrant features that show statistically significant different values between the pre- and post-migration periods, reflecting behavioral changes. After that, to discern between behavioral changes due to the new physical environment or to the migration process itself, we compare the post-migration migrant features against the same features computed for locals. Statistical analyses between the two distributions will allow us to identify behavioral differences that could be attributed as consequences of the internal migration movement.

6.1 Spatial Dynamics

We consider the following features to characterize individual spatial dynamics in terms of spatial dispersion, diversity and entropy:

1) *Number of Municipalities Visited.* Computed as the monthly average number of municipalities visited by an individual. This feature is calculated identifying the municipalities that correspond to the cellular towers where a given individual has been observed. As such, it will assess whether migrants are more prone to visit more municipalities than locals. This feature could be of interest to migration researchers interested in understanding trips to visit the social connections migrants left behind prior to migration.

2) *Entropy of Visits to Municipalities.* This feature measures the regularity of the visits to different municipalities. Insights from this feature could help future migration research focused on adaptation of migrants to new settings and routines.

3) *Daily mobility.* This feature is computed using the sequence of BTS towers visited in a day, which is an approximation of the real distance traveled by an individual. We focus on weekday traveled distances to assess the spatial dispersion of migrants potentially due to work-related reasons [11]. As opposed to municipality features that focus on long trips, daily mobility offers a window into short trip behaviors.

4) *Radius of gyration (ROG).* ROG is computed as the average area covered by of all the BTS towers used by an individual, weighted by the number of calls in each tower. ROG has been used in the literature as an approximation of the distance between home and work, which are the places where people spend most of their time, as has already been shown in the literature using CDR data [23]. We explore this feature as a proxy to evaluate commuting distances, which could motivate future migration researchers to understand the job dispersion that internal migrants are typically exposed to in Mexico. We calculate the ROG for each individual as the root mean square of the average distances from all other locations to his center of gravity (COG): $ROG_i = \sqrt{\frac{1}{N_i}\sum_{t=1}^{N_i}(d_{it} - COG_i)^2}$.

Finally, given that the spatial dynamics of an individual can be affected by the size and shape of the municipality, we normalize the daily mobility and ROG features by the radius of the municipality where the feature is computed. This process generates two additional spatial dynamics features: 5) Normalized daily mobility and 6) Normalized ROG. We compute all these spatial dynamic features for each individual in our dataset. Each migrant is characterized with a pre- and a post-migration distribution for each feature, and each local is characterized with a unique distribution per feature.

Next, we compute, for each spatial feature, a within-subjects t-test between the pre- and post-migration distributions to assess whether migrants change their spatial dynamics immediately after migration. Most of the tests were statistically significant at $p < 0.05$ *i.e.,* migrants change their spatial behaviors immediately post-migration. In this section, we focus on understanding whether these changes are due to the new physical environment or rather are consequences of the migration process itself. For that purpose, we compare migrants' post-migration behaviors against local behaviors to assess potential spatial dynamic consequences of the migrations. Specifically, we build, for each spatial feature, two population distributions representing the behavior of all locals in our sample and the behavior of all migrants in our sample for that feature. For each feature, we then compare the two distributions using Welch's t-test [66]. Additionally, to quantify the statistical differences between the two populations (effect size) we also compute Cohen's d [15]. The distributions of daily mobility and ROG can be heavily skewed by a small number of people who have extremely large distances traveled. To avoid the dominance influence of the outliers over the test result, we apply the Box-Cox transformation [51] to the data so that the population distribution is rendered close to normal. Table 1 shows the results for the statistical tests. Next, we describe our main observations.

Observation 1. We observe that in the immediate post-migration period, internal migrants in Mexico tend to visit more municipalities and have more irregular behaviors than the local community. The average number of municipalities migrants visit in a month is 2.0, compared to 1.6 for locals. The t-test shows that the difference is significant (t=34.38, p<$2.2e-16$), with a small effect size (Cohen's d value is 0.3189). A similar result was observed for the entropy of municipalities visited, with a medium effect size. In fact, immediately after migrating, migrants have, on average, higher entropy than locals, showing more irregular mobility patterns. Migrants were observed to visit both their pre-migration municipalities as well as municipalities located close to their post-migration destinations. We hypothesize that these findings could reveal that individuals make an effort to maintain their local connections in their pre-migration municipalities either because of work or personal reasons; in fact,

Features	Mean and SD after Box-Cox (locals / migrants)	Welch's t-test		Cohen's d	
		t	p-value	d	Effect Size
#1. Number of Municipalities	mean=1.66 sd=1.03 mean=2.00 sd=1.29	34.3860	<2.2e-16	0.3189	small
#2. Entropy of Municipalities	mean=0.14 sd=0.22 mean=0.28 sd=0.33	54.4800	<2.2e-16	0.5654	medium
#3. Daily Mobility *Box-Cox λ=0.1414*	mean=1.17 sd=0.22 mean=1.21 sd=0.24	16.2850	<2.2e-16	0.1480	small
#4. Normalized Daily Mobility *Box-Cox λ=0.1414*	mean=0.93 sd=0.19 mean=0.98 sd=0.21	28.4030	<2.2e-16	0.2551	small
#5. ROG *Box-Cox λ=0.1818*	mean=1.31 sd=0.40 mean=1.41 sd=0.59	21.6560	<2.2e-16	0.2312	small
#6. Normalized ROG *Box-Cox λ=0.1818*	mean=0.98 sd=0.32 mean=1.08 sd=0.45	28.7070	<2.2e-16	0.2961	small

Table 1: Statistical analysis of spatial dynamics' features: comparison between migrant and local behaviors.

similar findings have been qualitatively reported in other countries [14, 43]. Further qualitative analyses via interviews or surveys would be necessary to confirm or refute our hypothesis for Mexico. Due to the limitations of the temporal range of the data, we are not able to explore whether, in the long term, these behaviors remain more entropic o stabilize to levels similar to the local population.

Observation 2. The Table also shows that, immediately after migrating, migrants have significantly longer trips than locals (daily average mobility of 1.21 vs 1.17 after Box-Cox, 3.78km vs 3.12km prior to Box-Cox, with t=16.285, p<$2.2e-16$)). After a Box-Cox transformation to mitigate for outliers, and after spatial normalization to eliminate the role of the municipality size, the significant difference still prevails, with a small Cohen's d effect size. This result highlights that, at least during the first post-migration months, migrants appear to travel longer distances on a daily basis, which could be potentially related to: (i) migrants having more daily short trips than locals within a small geographical area or (ii) migrants having a unique, longer than locals, daily trip. We hypothesize that the former could be indicative of a larger informal job market within the migrant community [44]; while the latter could reveal larger job spatial dispersion (jobs are farther away for migrants than for locals), as shown in [3]. Further qualitative studies will be necessary to evaluate both hypotheses for Mexico.

Observation 3. As explained earlier, the ROG represents the geographical area or physical space where individuals spend *a vast majority* of their time. We make use of this variable to characterize the average distance between *home* and *work* [23]. Our analyses show that both ROG and normalized ROG are significantly different with a moderate difference in quantity between the two (small Cohen's d effect size). This result highlights that the differences in daily mobility discussed in *Observation 2* could be potentially due to longer commute trips, at least during the first months post-migration. Similar findings with respect to commuting distances were revealed by Browder *et al.* in an analysis of commuting patterns of internal migrants in Bangkok, Jakarta and Santiago using survey and interview data [8]. However, only further qualitative studies in Mexico will be able to clarify this hypothesis. As we have shown, the proposed framework can be used as a tool to reveal behavioral insights at a large scale that might motivate new qualitative studies for sociologists, geographers or ethnographers. For example, our analyses favored the longer commute trip hypothesis versus the informal market opportunities hypothesis. Understanding the reasons behind this finding, or whether it can also be replicated across countries, would be an interesting research question to analyze.

6.2 Social Ties

We consider the following features to model the online, cell phone-based social ties of migrants and locals:

1) *Number of Contacts per Month.* We compute the average number of monthly contacts the migrant talks to. We define contacts as individuals with whom reciprocal relationships are set up *i.e.,* where two individuals have at least one reciprocated phone call with a duration longer than five seconds. These filters are set up following Onnela's *et al.* work who have argued that reciprocal calls with long duration can be an indication of some work-, family-, leisure- or service- based relationship, while a single call every now and then may carry little information [48]. This feature will allow us to model the size of a migrant's cell phone-based social network, and compare it against average sizes of local networks.

2) *Number of Calls per Month.* This feature is used to understand the strength of the relationships that the migrant establishes with her cell phone-based social contacts.

3) *Entropy of Contacts.* This feature is computed weighting the cell phone-based social activity with each contact by its call frequency. The objective is to understand whether the migrant has a predictable communication pattern or whether it is more entropic.

4) *Number of New Contacts per Month.* We use the first three months of CDR data, prior to migration, to build the pre-migration, cell phone-based social network of a migrant. Onnela *et al.* showed that three consecutive months are enough to reconstruct the social network of an individual [48]. Next, during the post-migration period, we identify as new contacts those who share cell phone activity with an individual and who are not present in her pre-migration social network. This feature is critical to evaluate the temporal expansion of the cell phone-based social network in the post-migration period and quantify its growth.

5) To further characterize the cell phone-based social relationships of migrants, we build upon the feature of new contacts and analyze three additional features: 5a) *Ratio of (%) new contacts to the total number of contacts*, 5b) *Number of calls with new contacts*, and

5c) *Ratio of (%) calls with new contacts to all contacts*. Each of these features are computed as monthly average values per individual.

6) To assess the role that place plays in the cell phone-based social network of migrants, we evaluate: 6a) *Ratio of (%) local calls* defined as communications with contacts who live in the post-migration municipality with respect to all calls. The objective of this feature is to understand the weight that individuals give to communications with contacts from their previous home (pre-migration municipality) as opposed to their current home location (post-migration municipality); and 6b) *Ratio of (%) local calls with new contacts* with respect to all local calls, to analyze whether migrants mostly develop new cell phone-based contacts locally, in their new municipality, or in the distance, with their pre-migration municipality.

Similarly to the analysis with spatial dynamics features, we first evaluate whether pre- and post-migration social features are statistically significantly different and then compare post-migration behaviors against locals. The post-migration behaviors were statistically significantly different from their pre-migration behaviors via within-subjects t-tests ($p < 0.05$). We could not confirm significant differences for the new contact features since we do not have data prior to the pre-migration period to construct the social networks and identify the creation of new contacts in that period. Nevertheless, we also compare these features against the local behavior.

In this section, we focus on analyzing the behavioral differences between migrant and local populations with respect to social network features with the focus of forming potential hypotheses about the social consequences of the migration process. For that purpose, we compute the social features for all locals in our dataset, build the local population distributions for each feature, and compare them against the migrants' post-migration distributions using Welch's t-test and Cohen's d. Table 2 shows our results. Next, we discuss our main observations.

Observation 1. Immediately post migration, migrants communicate with a similar volume of calls than locals, but with a slightly smaller number of contacts. The Table shows that the difference between migrants and locals is statistically significant, with a small Cohen's d effect size for the number of contacts, and negligible for the number of calls. This results shows that, in the large scale, internal migrants appear to have fewer social connections, and as a result, more frequent communications with their contacts than their local counterpart. In other words, immediately after migration, internal migrants rely on stronger, cell phone-based social relationships than locals.

Observation 2. In the immediate post-migration period, migrants show lower entropy in their cell phone-based social networks than locals (mean=1.12, sd=0.64 vs. mean=1.32, sd=0.63). This highlights the fact that, during their first months post-migration, migrants appear to have a more regular calling behavior. Given that the percentage of local calls is significantly smaller for migrants (see #8 in Table 2) we hypothesize that migrants maintain regular calling patterns with their pre-migration municipalities, potentially due to work or family. Similar results have been shown in the context of international migration [60], however further qualitative studies would be necessary to confirm the hypothesis for internal migration in Mexico.

Observation 3. In the first three months after migration, migrants tend to add fewer monthly contacts to their social network than locals. In fact, the ratio of new contacts to the existing social network shows a small statistical difference with a mean of 7% for migrants and 11% for locals. This result shows that immediately after migrating, migrants appear to have more difficulties in expanding their cell phone-based social networks than non-migrants. In fact, migrants appear to favor tight communications with pre-migration contacts since the ratio of local calls shows significantly lower numbers for communications with local contacts. We hypothesize that this observation might be supported by the theory of acculturation stating that migrants usually have difficulties in their adaption process [37]. While migrants might have the necessity to re-build their disrupted social network, they have difficulties to expand it when compared to non-migrant behaviors. However, further qualitative research would be necessary to confirm such theory in the context on internal migration in Mexico.

Observation 4. Immediately after migration, migrants have fewer communications with locals than the locals themselves. In fact, while only 37% of the calls that migrants make are with locals, 73% of the calls from locals are to other locals (non-migrants). A similar behavior is observed with the new contacts made. We observe that the ratio of local calls with new contacts is also significantly smaller for migrants; with only 46% of the migrants' calling behavior taking place with new local contacts (as opposed to 74% for locals), which reflects that migrants also continue to expand their social network in their pre-migration communities. These findings reflect that the first months after migration, migrants still heavily rely on their pre-migration social network; we hypothesize that these findings could be explained with the social support theory that determines that migrants tend to seek for friends and family social support to buffer the negative effects of migration stress [12]. However, further qualitative analysis would be necessary to confirm or refute such theory for Mexican migration.

7 ROLE OF PRE-MIGRATION BEHAVIORS ON POST-MIGRATION ACTIVITIES

We analyze the role that pre-migration spatial and social features might play in the behavioral changes observed immediately post-migration as migrants adapt to their new communities. For that purpose, we run multivariate regression models on the pre-migration features, with a focus on analyzing (i) the relationship between pre- and post-migration behaviors, (ii) the statistical significance and importance of pre-migration features with respect to the post-migration behaviors, and (iii) to what extent such models can be used to predict post-migration behaviors given pre-migration information. We compute multivariate regression models with each spatial dynamics and social network feature described in the previous section as both dependent and independent variables. The models are built on a subset of those variables *i.e.,* the independent variables do not include social network features with respect to new contacts since we do not have data prior to the pre-migration period to construct the social networks and identify the creation of new contacts.

We have tested for multicollinearity between the independent variables used in the regressions via both the variance inflation factor (VIF) and the condition index (CI) [52]. The largest VIF and CI values were for the variable *number of contacts* (with $VIF = 4.3$ and

Features	Mean and SD after Box-Cox locals migrants	Welch's T-test		Cohen's d	
		t	p-value	d	Effect Size
#1. Number of Contacts per Month *Box-Cox* $\lambda = 0.1414$	mean=1.31 sd=0.13 mean=1.29 sd=0.13	-27.7800	<2.2e-16	-0.2145	small
#2. Number of Calls per Month *Box-Cox* $\lambda = -0.1414$	mean=1.87 sd=0.26 mean=1.86 sd=0.29	-4.0602	4.92e-5	-0.0344	negligible
#3. Entropy of Contacts	mean=1.32 sd=0.63 mean=1.12 sd=0.64	-40.658	<2.2e-16	-0.3205	small
#4. Number of New Contacts per Month *Box-Cox* λ*=-0.1010*	mean=1.02 sd=0.09 mean=0.99 sd=0.07	-29.7430	<2.2e-16	-0.2526	small
#5. % New Contacts	mean=0.11 sd=0.10 mean=0.07 sd=0.13	-39.107	<2.2e-16	-0.3935	small
#6. Number of Calls with New Contacts *Box-Cox* λ*=0.0202*	mean=1.03 sd=0.03 mean=1.04 sd=0.03	19.514	<2.2e-16	-0.1906	small
#7. % Calls with New Contacts *Box-Cox* λ*=0.1818*	mean=0.60 sd=0.12 mean=0.62 sd=0.14	14.2720	<2.2e-16	-0.1600	small
#8. % Local Calls	mean=0.73 sd=0.32 mean=0.37 sd=0.37	-117.0900	<2.2e-16	-1.0871	large
#9. % Local Calls with New Contacts	mean=0.74 sd=0.38 mean=0.46 sd=0.45	-43.6890	<2.2e-16	-0.7204	medium

Table 2: Statistical analysis of social ties' features: comparison between migrant and local behaviors using Welch's t-test with Cohen's d and Box-Cox transformation for skewed distributions.

$CI = 12$), while other values were in the range of $1.3 < VIF < 3.5$ and $1 < CI < 7$. Although $VIF < 10$ and $CI < 30$ are considered acceptable [28], we evaluated multivariate regression models with and without the *number of contacts* variable. Table 3 shows the results for each model (one model per line) taking into account the number of contacts. Removing that variable due to a potentially low multicollinearity changed minimally the coefficients (size), but the sign and significance were the same and thus are not reported in the paper. Next, we discuss our main findings.

Observation 1. Post-migration spatial and social features are highly influenced by their pre-migration values *i.e.,* one of the most predictive features for any given post-migration spatial or social feature is its own pre-migration value. This is reflected by the significant and positive coefficients (at p<0.001) between the same pre- and post-migration features, with the exception of the normalized ROG (see bold coefficients in the diagonal of Table 3).

However, looking at the value of the regression coefficients, we can observe that the pre-migration features impact differently their post-migration values. For spatial dynamic features (ROG) the impact is small *i.e.,* we have a small rate of change for the post-migration features when its pre-migration values change one unit and all other pre-migration features are kept the same. On the other hand, the social network features have a larger impact, with rates of change per pre-migration unit having a larger impact on its post-migration values. For example, the radius of gyration before migration (ROG, line #1 in Table) can lead to 0.1713 (***) times change in its post-migration value if the other independent features are kept fixed; while for the number of monthly contacts (line #3) the rate of change in the post-migration values with respect to their pre-migration ones is large: 0.7821 (***). These results also highlight that social network features typically experience larger quantitative changes between the pre- and post-migration stages than the spatial dynamic features, whose changes between pre- and post-migration stages are more modest.

Observation 2. The cell phone-based social relationships that migrants have in the pre-migration period are highly indicative of the types of cell phone-based relationships they will have in their immediate post-migration lives. We can observe that the ratio of calls made to the post-migration municipality prior to migrating is significantly and positively related to the post-migration ratio of local calls (#6) and to the post-migration ratio of local calls with new contacts (#7) *i.e.,* migrants who start developing new cell phone-based social relationships with their post-migration municipalities even before they migrate, are prone to have more local communications and more new local contacts once they move to their post-migration destination.

Similarly, the number of contacts pre-migration is also significantly related (0.0697 at significance level 0.05) to the number of new contacts post-migration (#8) *i.e.,* people who are more socially connected prior to migrating will highly probably develop a large social network with new contacts in their post-migration communities, immediately after migrating. Observations 1 and 2 could potentially motivate further qualitative studies to assess the theory that migrants' personality characteristics can partially explain their adaptability in the host society [37]. Although it is not enough to fully explain how they cope with their new environment, previous work has shown that people who communicate more are generally better at re-rooting their social network in the host society.

Observation 3. The pre-migration ratio of calls made to the post-migration municipality is statistically significantly related to both the ROG (#1) and the normalized ROG (#2). In fact, we observe a significant and high negative coefficient for the ROG (−18.5053 ***) and for the normalized ROG (−6.3999 ***). This result highlights that having social connections in the pre-migration period with

Post-migration Features	Pre-migration Features						Intercept	Adjusted R-squared	p-value
	ROG	Normalized ROG	Number of Contacts	Number of Calls	Entropy of Contacts	% Calls to Post-migration Home			
#1. ROG	**0.1713*****	-0.0378	0.9931***	-0.0107*	-1.7578	-18.5053***	24.3852***	0.0281	<2.2e-16
#2. Normalized ROG	0.0334***	**-0.0062**	0.0992	0.0013	0.1497	-6.3999***	7.0510***	*0.0120*	<2.2e-16
#3. Number of Contacts	0.0003	0.0003	**0.7821*****	-0.0012***	-0.3805***	-0.2751*	2.7170***	**0.4723**	<2.2e-16
#4. Number of Calls	-0.0365	-0.0200	5.0428***	**0.5388*****	-27.8281***	-32.8129***	70.9551***	0.264	<2.2e-16
#5. Entropy of Contacts	0.0001	0.0001	0.0240***	-0.0003***	**0.4415*****	0.0450**	0.4961***	**0.3516**	<2.2e-16
#6. % Local Calls	0.0003***	0.0000	0.0009	-6.854e-5***	0.0258**	**0.5942*****	0.1643***	**0.2270**	<2.2e-16
#7. % Local Calls with New Contacts	0.0003**	0.0010*	0.0020	-0.0001.	0.0019	**0.2963*****	0.3562***	*0.0461*	<2.2e-16
#8. Number of New Contacts	0.0001	0.0003	**0.0795*****	-0.0004****	-0.0918**	-0.0564.	0.4421***	0.0926	<2.2e-16

Significance levels: 0 ’***’; 0.001 ’**’; 0.01 ’*’; 0.05 ’.’; 0.1 ’ ’

Table 3: Multivariate regressions on pre-migration behavioral features to quantify their role on post-migration behaviors. Adjusted R-squared values are used to assess the predictability of post-migration behaviors with pre-migration features.

the post-migration destination will affect the estimated rate of change for both spatial dynamics variables by a high negative factor when all other features are kept fixed. In other words, the more social connections during pre-migration with the post-migration municipality, the larger the decrease in the average length of the commuting patterns in the immediate post-migration period.

This result is really interesting when combined with a result obtained in the previous Section 6. The analysis comparing post-migration behaviors between migrants and locals showed that, on average, migrants tend to have larger ROGs and normalized ROGs. The current analysis appears to be showing that, if the migrant already has a connection with the post-migration destination, the ROG will be smaller, and potentially closer to that of locals. Further qualitative studies could look into whether this result could be interpreted as an indication of migrants adapting more quickly to their new environments when they previously have social connections with and knowledge about those environments. Similar results have been shown by Holmes *et al.* who interviewed migrants and assessed that communities with access to information from their post-migration communities prior to migration helped in bridging the challenges of adaptation [30].

Observation 4. The adjusted R-squared values shown in the Table measure how much variance of the dependent variables (the post-migration behavioral features) can be explained by pre-migration features. Generally, these values are good for some of the social network features ($0.264 < R^2 < 0.4723$), and poor for the spatial dynamics features (maximum adjusted $R^2 = 0.0281$). Specifically, the regression models for the number of contacts, the entropy of contacts and the ratio of local calls are relatively high, showing that these post-migration behaviors can be partially predicted with the subset of pre-migration features considered. However, the models for ROG, normalized ROG and the ratio of new contacts have small R-squared values, revealing that we can not use only pre-migration features to predict these post-migration values.

8 LIMITATIONS

At the time when the CDR data was processed (2009 – 2010), it is estimated that the percentage of cell phone owners in Mexico was approximately 60% [26]. Additionally, given that we only have access to CDR data from one cell phone company, this limits the representativeness of the local and migrant populations. Previous work has shown that individuals in CDR data from a country in Latin America represented a population with different demographic and socioeconomic characteristics, with percentages of cell phone users per strata similar to the actual population census [21]. However, in this paper we introduce yet another type of selection bias. Since we infer home location from CDR data, individuals that do not have call activity at night, are not assigned a home location and, as a result, are not part of the behavioral analysis. This approach implies that the behavior of people who have an overall reduced number of communications will probably be filtered out from the analyses; which might translate into loosing information from a specific migrant group associated to low cell phone use.

A second limitation of the work presented in this paper is the temporal length of the dataset used, which is of eight months. Since we focus our analyses on immediate post-migration behaviors only, in principle, we do not require long periods of CDR data after migration. However, due to the limited temporal length, our approach is not able to differentiate between long-term and short-term migrants. In any case, the methods and results presented in this paper are still highly valuable to understand immediate post-migration behaviors across typologies of internal migrants.

A third limitation of the paper is the CDR data itself. Although this data can reveal spatial and cell phone-based social behaviors of migrants, it lacks insight into the reasons why the observed behaviors take place including, but not limited to, the psychological and decision-making processes related to the migration experience. Finally, it is important to highlight that our analyses use data subject to privacy concerns. Although the data is anonymized and the results are reported in an aggregated manner, it has been previously shown that under some circumstances CDR data can be potentially used to uniquely identify individuals [17].

9 CONCLUSION

In this paper, we have proposed a framework that uses cell phone metadata to quantitatively analyze immediate post-migration behaviors and the role of pre-migration behaviors from two perspectives: the spatial dynamics and the social ties of migrants. We have revealed a large number of behavioral findings that we expect will guide or complement current research trends in micro-level migration studies working to understand the physical, social and psychological decision processes behind migration experiences. Ultimately, we envision this framework as part of a larger project that, combining quantitative and qualitative research, provides insights for the development of policies that help internal migrants in their adaptation process, with the objective of empowering them and reducing their vulnerability.

REFERENCES

[1] R. Ahas, S. Silm, O. Järv, E. Saluveer, and M. Tiru. 2010. Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of urban technology* 17, 1 (2010), 3–27.
[2] David Akeju. 2013. Africa, internal migration. *The Encyclopedia of Human Migration* (2013).
[3] Jeffrey J Axisa, K Bruce Newbold, and Darren M Scott. 2012. Migration, urban growth and commuting distance in Toronto's commuter shed. *Area* 44, 3 (2012), 344–355.
[4] Richard D Bedford et al. 2017. *New Hebridean Mobility: a study of circular migration.* Canberra, ACT: Dept. of Human Geography, Research School of Pacific Studies, The Australian National University.
[5] Joshua E Blumenstock. 2012. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. *Information Technology for Development* 18, 2 (2012), 107–125.
[6] A Bonfiglio. 2012. New approaches for researching the determinants of migration processes: ESF strategic workshop on migration research. *International Migration Institute, University of Oxford, for European Science Foundation (ESF)* (2012).
[7] George J Borjas, Stephen G Bronars, and Stephen J Trejo. 1992. Self-selection and internal migration in the United States. *Journal of urban Economics* 32, 2 (1992), 159–185.
[8] John O Browder, James R Bohland, and Joseph L Scarpaci. 1995. Patterns of development on the metropolitan fringe: Urban fringe expansion in Bangkok, Jakarta, and Santiago. *Journal of the American Planning Association* 61, 3 (1995), 310–327.
[9] LA Brown and J Holmes. 1971. Intra-urban migrant lifelines: a spatial view. *Demography* 8, 1 (1971), 103–122.
[10] Martin T Cadwallader. 1992. *Migration and residential mobility: Macro and micro approaches.* Univ of Wisconsin Press.
[11] Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2011. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. *IEEE Pervasive Computing* 10, 4 (2011), 36–44.
[12] Arul Chib, Holley A Wilkin, and Sri Ranjini Mei Hua. 2013. International migrant workersâĂŹ use of mobile phones to seek social support in Singapore. *Information Technologies & International Development* 9, 4 (2013), pp–19.
[13] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 1082–1090.
[14] Cati Coe. 2011. What is the impact of transnational migration on family life? Women's comparisons of internal and international migration in a small town in Ghana. *American Ethnologist* 38, 1 (2011), 148–163.
[15] Jacob Cohen. 1992. A power primer. *Psychological bulletin* 112, 1 (1992), 155.
[16] G.F. De Jong and J.T. Fawcett. 1981. Motivations for migration: an assessment and a value-expectancy research model. (1981).
[17] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3 (2013), 1376.
[18] Bruce Frayne and Wade Pendleton. 2001. Migration in Namibia: Combining macro and micro approaches to research design and analysis. *International Migration Review* 35, 4 (2001), 1054–1085.
[19] E. Frias-Martinez and V. Karamcheti. 2002. A customizable behavior model for temporal prediction of web user sequences. In *Int. Workshop on Mining Web Data for Discovering Usage Patterns and Profiles.* 66–85.
[20] Vanessa Frias-Martinez, Cristina Soguero-Ruiz, Enrique Frias-Martinez, and Malvina Josephidou. 2013. Forecasting socioeconomic trends with cell phone records. In *Proc.3rd ACM Symp. Computing for Development.* ACM, 15.
[21] Vanessa Frias-Martinez and Jesus Virseda. 2013. Cell phone analytics: Scaling human behavior studies into the millions. *Information Technologies & International Development* 9, 2 (2013), pp–35.
[22] V. Frias-Martinez, J. Virseda, A. Rubio, and E. Frias-Martinez. 2010. Towards large scale technology impact analyses: Automatic residential localization from mobile phone-call data. In *4th ACM/IEEE Intl. Conf. Inf. and Communication technologies and development.* ACM, 11.
[23] M. Gonzalez, C. Hidalgo, and A.-L. Barabasi. 2008. Understanding individual human mobility patterns. *nature* 453, 7196 (2008), 779.
[24] Nicola Green. 2002. On the move: Technology, mobility, and the mediation of social time and space. *The information society* 18, 4 (2002), 281–292.
[25] Michael J Greenwood. 1975. Research on internal migration in the United States: a survey. *Journal of Economic Literature* (1975), 397–433.
[26] GSMA. [n. d.]. GSMA Country Overview: Mexico, Mobile Driving Growth, Innovation and Opportunity. https://www.gsma.com/latinamerica/wp-content/uploads/2016/06/report-mexico2016-EN.pdf. Accessed: 04-10-2018.
[27] Douglas T Gurak and Fee Caces. 1992. Migration networks and the shaping of migration systems. *International migration systems: A global approach* (1992), 150–176.
[28] Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, Ronald L Tatham, et al. 2006. *Multivariate data analysis (Vol. 6).* Upper Saddle River, NJ: Pearson Prentice Hall.
[29] Martijn Hendriks, Kai Ludwigs, and Ruut Veenhoven. 2016. Why are locals happier than internal migrants? The role of daily life. *Social Indicators Research* 125, 2 (2016), 481–508.
[30] Prue Holmes and Annick Janson. 2008. Migrants' communication practices with ICTs: tools for facilitating migration and adaptation? *International journal of technology, knowledge & society.* 4, 6 (2008), 51–62.
[31] Emma Hopkins, Francesca Bastagli, and Jessica Hagen-Zanker. 2016. *Internal migrants and social protection: a review of eligibility and take-up.* Technical Report. ODI Working Paper.
[32] G. Hugo. 2013. What we know about circular migration and enhanced mobility. *Migration Policy Institute* 7 (2013).
[33] Mexican Statistical Institute INEGI. 2013. 2010 Census of Population and Housing Units. http://en.www.inegi.org.mx/proyectos/ccpv/2010/. Accessed: 2017-09-12.
[34] Mexican Statistical Institute INEGI. 2015. National Survey of Demographic Dynamics 2014. http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/enadid/2014/default.html. Accessed: 2017-06-15.
[35] S. Isaacman, R. Becker, R. Caceres, S. Kobourov, J. Martonosi, M.and Rowland, and A. Varshavsky. 2011. Identifying important places in peoples lives from cellular network data. In *Int. Conf. on Pervasive Computing.* 133–151.
[36] Ben CH Kuo. 2014. Coping, acculturation, and psychological adaptation among migrants: a theoretical and empirical review and synthesis of the literature. *Health Psychology and Behavioral Medicine* 2, 1 (2014), 16–33.
[37] Wen H Kuo and Yung-Mei Tsai. 1986. Social networking, hardiness and immigrant's mental health. *Journal of health and social behavior* (1986), 133–149.
[38] M. Lenormand, A. Bassolas, and J.J. Ramasco. 2016. Systematic comparison of trip distribution laws and models. *J. Transport Geography* 51 (2016), 158–169.
[39] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant. 2012. A universal model of commuting networks. *PloS one* 7, 10 (2012), e45985.
[40] Maxime Lenormand, Miguel Picornell, Oliva G Cantú-Ros, Thomas Louail, Ricardo Herranz, Marc Barthelemy, Enrique Frías-Martínez, Maxi San Miguel, and José J Ramasco. 2015. Comparing and modelling land use organization in cities. *Royal Society open science* 2, 12 (2015), 150449.
[41] Shengxiao Li and Pengjun Zhao. 2018. Restrained mobility in a high-accessible and migrant-rich area in downtown Beijing. *European transport research review* 10, 1 (2018), 4.
[42] Yanwei Lin, Qi Zhang, Wen Chen, Jingrong Shi, Siqi Han, Xiaolei Song, Yong Xu, and Li Ling. 2016. Association between social integration and health among internal migrants in ZhongShan, China. *PloS one* 11, 2 (2016), e0148397.
[43] Jennifer Mason. 2004. Managing kinship over long distances: the significance of âĂŸthe visitâĂŹ. *Social Policy and Society* 3, 4 (2004), 421–429.
[44] Dipak Mazumdar. 1976. The urban informal sector. *World development* 4, 8 (1976), 655–679.
[45] Raven Molloy, Christopher L Smith, and Abigail Wozniak. 2011. Internal migration in the United States. *Journal of Economic perspectives* 25, 3 (2011), 173–96.
[46] Raul Montoliu and Daniel Gatica-Perez. 2010. Discovering human places of interest from multimodal mobile phone data. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia.* ACM, 12.
[47] J.M. Nogle. 1994. Internal migration for recent immigrants to Canada. *International Migration Review* (1994), 31–48.
[48] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. 2007. Structure and tie strengths in mobile communication networks. *Proc. Nat. Academy of Sciences* 104, 18 (2007), 7332–7336.
[49] David A Plane. 1984. Migration space: Doubly constrained gravity model mapping of relative interstate separation. *Annals of the Association of American Geographers* 74, 2 (1984), 244–256.
[50] A. Rubio, V. Frias-Martinez, E. Frias-Martinez, and N. Oliver. 2010. Human Mobility in Advanced and Developing Economies: A Comparative Analysis.. In *AAAI Spring Symposium: Artificial Intelligence for Development.*
[51] RM Sakia. 1992. The Box-Cox transformation technique: a review. *The statistician* (1992), 169–178.
[52] George AF Seber and Alan J Lee. 2012. *Linear regression analysis.* Vol. 329. John Wiley & Sons.
[53] F. Simini, M.C. González, A. Maritan, and A.lbert-László Barabási. 2012. A universal model for mobility and migration patterns. *Nature* 484, 7392 (2012), 96.
[54] Ronald Skeldon. 2014. *Migration and development: A global perspective.* Routledge.
[55] Ronald Skeldon. 2017. International Migration, internal migration, mobility and urbanization:towards more integrated approaches. *Population Division, Department of Economic and Social Affairs, United Nations* (2017).
[56] S.A. Stouffer. 1940. Intervening opportunities: a theory relating mobility and distance. *American sociological review* 5, 6 (1940), 845–867.
[57] Waldo Tobler. 1995. Migration: Ravenstein, thornthwaite, and beyond. *Urban Geography* 16, 4 (1995), 327–343.

[58] Michael Todaro. 1980. Internal migration in developing countries: a survey. In *Population and economic change in developing countries*. University of Chicago Press, 361–402.

[59] Zeynep Tufekci. 2014. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *ICWSM* 14 (2014), 505–514.

[60] Steven Vertovec. 2004. Cheap calls: the social glue of migrant transnationalism. *Global networks* 4, 2 (2004), 219–224.

[61] Ingmar Weber, Emilio Zagheni, et al. 2013. Studying inter-national mobility through IP geolocation. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 265–274.

[62] Yingxiang Yang, Carlos Herrera, Nathan Eagle, and Marta C González. 2014. Limits of predictability in commuting flows in the absence of data for calibration. *Scientific reports* 4 (2014), 5662.

[63] Emilio Zagheni, Venkata Rama Kiran Garimella, Ingmar Weber, et al. 2014. Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 439–444.

[64] Emilio Zagheni and Ingmar Weber. 2012. You are where you e-mail: using e-mail data to estimate international migration rates. In *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 348–351.

[65] Cathy Zimmerman, Ligia Kiss, and Mazeda Hossain. 2011. Migration and health: a framework for 21st century policy-making. *PLOS medicine* 8, 5 (2011), e1001034.

[66] Donald W Zimmerman and Bruno D Zumbo. 1993. Rank transformations and the power of the Student t test and Welch t'test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 47, 3 (1993), 523.

[67] G.K. Zipf. 1946. On the intercity movement of persons. *American sociological review* 11, 6 (1946), 677–686.