

A Mobility Model for Return and Repeated Migration Based on Network Motifs

Rafael Prieto Curiel (✉ rafael.prieto.13@ucl.ac.uk)

University College London

Mauricio Quiñones Domínguez

EAFIT University

Eduardo Lora

EAFIT University

Neave O'Clery

University College London

Research Article

Keywords: Network, Migrations, distinct locations

Posted Date: July 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-655935/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A mobility model for return and repeated migration based on network motifs

Rafael Prieto Curiel^{1*} Mauricio Quiñones Domínguez²
Eduardo Lora²
Neave O'Clery¹

1 Centre for Advanced Spatial Analysis, University College London,
Gower Street, WC1E 6BT, London, United Kingdom

2 Research in Spatial Economics, EAFIT, Carrera 49 N° 7 Sur-50
Medellín, Colombia

* corresponding author: rafael.prieto.13@ucl.ac.uk

Abstract

People changing residence within their own country is the most common type of migration and is one of the main driving forces of a country's demography. Yet, due to a lack of data and tools, some key characteristics of internal migration such as the propensity to remain in the same city or to return to previous locations are frequently ignored. Here, a network-based model of city-to-city migration is constructed, where the movement of individuals is modelled using the frequency of distinct network motifs. We fit this model to longitudinal data on 3.3 million workers in Colombia, including 1.4 million migrations, and compare the motif frequency based on migration and return rates between men and women, and between distinct age and income groups. Results show that the majority of people do not move in general, but that there is a small group which exhibits frequent migration, particularly the young and male. Nearly three out of four times that a person moves, they are going back to a previous city, although women and mature people are less likely to move and more likely to return if they actually move. At a city level, different onward and return migration patterns are observed, whereby people from small secondary towns are more likely to leave and not return than people from large metropolitan areas like Bogotá or Medellín.

1 Introduction

Our understanding of human mobility has drastically evolved over the past few decades, mainly due to the availability of geolocated datasets (1; 2; 3) combined with processing power. A specific type of mobility characterised by long distances and periods of time is migration (4; 5) which has an impact at a city, country and global level (6). Migration is the principal source of population redistribution and changes in the patterns of human settlement (7; 8). Further, it tends to accelerate ageing in places where the young are more likely to move, increasing the pressure on those who remain, and deepens the gender imbalance by altering the gender ratio of a place (9; 10; 11).

Internal migration is a key driver of metropolitan development in many parts of the World (12) and the sorting mechanism of productive workers across cities (13). In the

US, for instance, more than 8 million people move each year from one metropolitan area to another, whilst in the UK, 10% of the population change residence each year (14). Similar patterns are observed in other countries. For example, migration between urban areas accounted for more than half of all internal population flows in Panama, Paraguay and Brazil (15). Individual decisions of those who decide to move or stay collectively determine the macro behaviour at neighbourhood, city and country level (16) and so models of human migration are a primary tool for forecasting city size and its demand for resources (17; 18; 19).

Frequently, when a person migrates, they move back to places in which they lived before. For many reasons, perhaps due to unemployment after moving or to reduce the emotional or monetary costs of being far away from home, people are frequently return migrants (20; 13). Hence, some movements can be classified as onwards migration, while others are classified as return migration, if she or he goes back to any location in which they previously lived. There is a limited literature on return migration. Notably, and perhaps unsurprisingly, the longer a person lives in location, the lower probability that the person will leave their city (21; 22). Cumulative inertia is observed both for people who move and for people who decide to stay in a place. Only a small population group is ever a migrant. Hence we might observe those who frequently keep moving, called the "repeat migrants", and those who never move, called the "stayers" (20). Migration is a path-dependent complex phenomena, whereby only a few individuals move frequently, often moving to previously-lived places. Yet, the level of concentration of migration and the rate at which different population groups return has not yet been fully captured due to the complexity of return and repeated movement patterns and the difficulty of measuring these patterns in real-world data.

The interest in understanding generalisable patterns of migration dates back decades. The laws of migration (23), published by Ernst Georg Ravenstein in 1885, were developed by looking at migration data at a county level from and to the UK, Ireland and Scotland. The law states, among other things, that the majority of migrants move short distances (24) and large towns grow as a result of migration rather than natural population growth (6). Since then, a variety of novel data sources that help capture more aspects of migration (beyond census data) have been used to analyse migration. These include, for example, data from social media (25), card transactions (26), mobile phone data (27; 28; 29), and others (30; 31; 32). These studies have shown, for example, that women tend to make shorter trips (33) and visit less diverse places than men (34). Using mobile phone call records and manual data collection, researchers analysed daily trip patterns to classify individuals as explorers and returners (30), and showed that daily trips can be described by a limited number of network patterns (35). Nevertheless, in terms of migration, a classification of explorers and returners within a framework that captures distinct types of complex movements is missing.

One of the most significant challenges for the study of migration is that mobility data is often a byproduct of sources which were not originally intended to provide information on the movements of people (36). Significant details are often unknown. For instance, individuals are frequently observed from an arbitrary starting point (particularly when individuals are not surveyed), which means that their precise origin (and likely first migrations) are ignored. Thus, it is impossible to detect if a person moves frequently or if the person is returning, as previous locations are not captured by the data. Many migration studies are based on surveys, with different waves of individuals each time, and so migration is treated as a one-time event (37). Hence, in most cases

path-dependency in terms of migration is difficult to capture. Detecting the extent to which individuals tend to migrate and return between cities, and whether gender, age or income play a role in an individual’s return frequency, is critical to better understand internal migration patterns and to characterise distinct types of migrants.

Here we construct a model to capture the complex patterns formed by people who move. The model has two key components. First, we represent migration signatures via motifs on a network (38; 39). This approach enables us to measure and compare the path-dependent intensity of return and onward migration. Network motifs are frequently used in biology, ecology, engineering (39), in social network analysis (40), and even to detect crime patterns (41), as well as mobility patterns (35). Second, we develop a probabilistic model in order to transform the raw data of migration counts into migration rates which explicitly take into account the fact that each person is observed only for a finite window of time. Hence, this approach means that we can ignore whether we are observing the first or subsequent migrations of an individual. We then apply a mixture model to these rates to estimate the number and size of population groups corresponding to distinct rates. This model fits two model parameters to the observed motifs, the migration intensity and the return propensity, and enables us to reproduce complex migration patterns and measure the gender, age and income divide in terms of migration.

Latin America is one of the most urbanised regions in the world, moving from just 30% urbanised in 1950 to over 85% today. With more than 80% of its 50 million inhabitants living in one of 62 geographically isolated cities, Colombia is representative of highly urbanised middle-income Latin American nations and presents an ideal observatory from which to observe and analyse inter-city migration patterns. In this paper we use data from the Colombian Social Security (PILA) which contains nine consecutive years of administrative data for 3.3 million formal employees between 2008 and 2016. Within this time period we extract 1.4 million city-to-city migrations.

2 Results

2.1 Representing onward and return migration

Our main data source is administrative data on formal employees in Colombia provided by firms to the Ministry of Health and Social Protection. The Planilla Integrada de Liquidación de Aportes (PILA) dataset we use spans 2008-16 and contains 3.3 million individuals. As detailed in the Methods section, we use this dataset to detect 1.4 million inter-city migrations. We note that this data does not capture international migration or internally displaced people due to conflict, which are combined thought to have involved over 10 million people (42).

The frequency of specific patterns of onward and return migration is captured by constructing migration motifs for all individuals on a network. Mobility networks have been used to study migration (43; 44; 20) and daily human mobility (45; 35). These studies found that most mobility patterns can be described by a small number of daily movements or “motifs” (2). Schematically (as in (37)), we represent the known locations of an individual across the 9 years we observe them in our dataset as a sequence of 9 characters. For example, we could have $AAA|AAA|AAA$, $AAA|AAB|BBB$, or $AAA|ABB|CCA$, where A , B and C represent different cities (or the countryside which is also represented by a single letter in our case). Hence, instead of a labelled network

where each node corresponds to a distinct city, we construct a network where all individuals start at node A , and if they move to new locations, they move in alphabetical order (so they move to node B , then C and so on). If a person moves for the second time, then return migration forms the sequence ABA and onward migration forms the sequence ABC . We then use this sequence of locations to construct motifs, which correspond to this sequence but removing repetitions of consecutive locations. Hence, if an individual does not move, its motif is simply A . If an individual moves once, it has motif AB . If a person moves three times and the last one is a return migration, then the motif would be $ABCA$. Some further examples of motifs are shown in Table 1.

Known locations	Motif	Migrations M_i	Locations L_i	Number of returns	Number of onward
$AAA ABB BBB$	AB	1	2	0	1
$AAB BBB BAA$	ABA	2	2	1	1
$AAA ABB CAA$	$ABCA$	3	3	1	2
$AAB ACC CCB$	$ABACB$	4	3	2	2
$ABC CDD BAC$	$ABCD BAC$	6	4	3	3

Table 1: Number of migrations, distinct locations, return migrations and onward migrations based on the person’s imputed data.

Many individuals do not move during the nine years of the data, but some move more than once according to different motifs. In particular, complex motifs which include a return to a previous city are very frequent. For example, from the group of individuals who move twice (8% of the people), it is observed that 81% of them returned to their previous location and only 19% exhibited onwards migration. And individuals who moved three times are 2.9 times more likely to have moved only between two cities than between four cities. Overall, we find that 95% of individuals who move can be described with motifs which include just three three distinct characters, such as AB , $ABAC$, $ABCAB$, $ABABC$. Thus, there is a very high rate of return migration (Figure 1). Other, more complicated motifs are also observed, for example, the motif $ABCADA$ was observed among 0.19% of individuals. See the Appendix for the frequency of the top 30 motifs.

Below we propose a parameter-based model to generate a similar frequency of distinct motifs as the observed ones. We then fit these parameters to the data in order to empirically capture the return rates and the concentration of migration for particular populations.

2.2 A migration model based on a network

One of the problems with data collection in a passive manner is that we begin observing a person at some arbitrary starting point. This means that we do not know if their first known location corresponds to their origin, or even if their first (known) migration is a return migration to a city in which they previously lived or an onward migration. Instead of assuming that people have not moved before we begin observing them, we propose a method to estimate the migration rate of each person and then group individuals based on their migration rate. This technique has been applied to crime data, for example, where people are usually observed via yearly victimisation surveys.

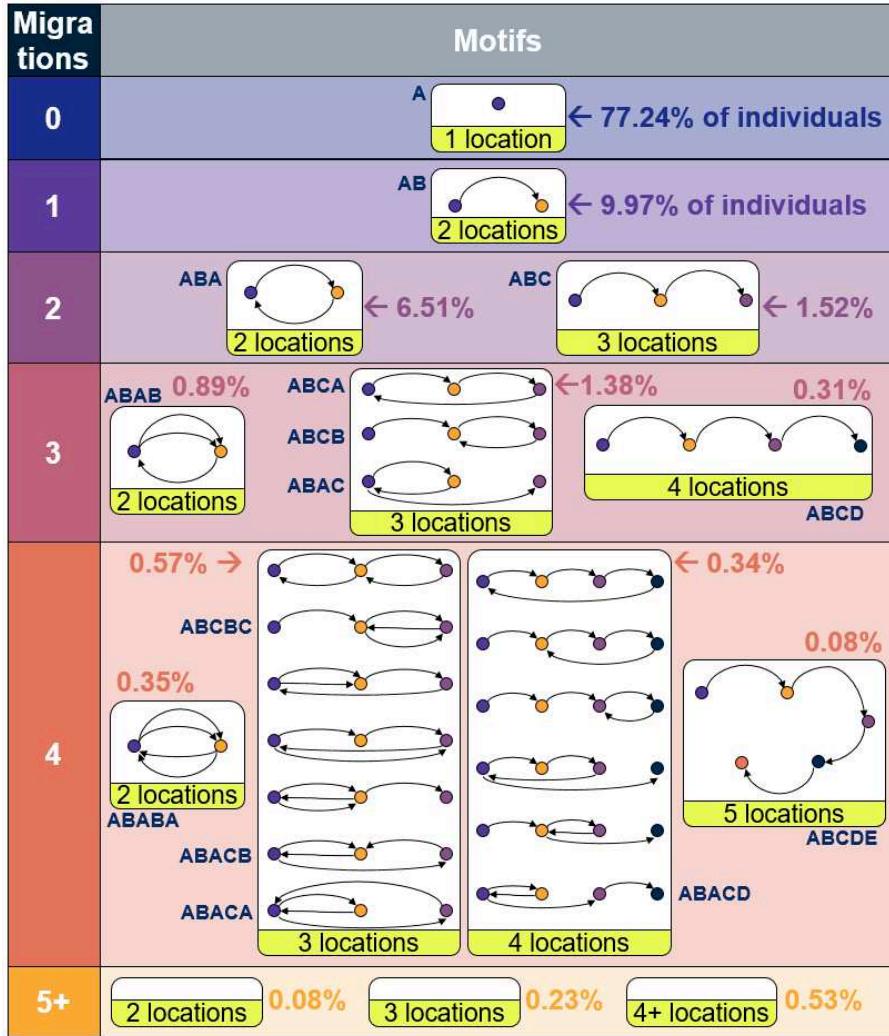


Figure 1: Distribution of motifs when the person moves four times or less. All individuals begin in the blue node (left) and move according to the arrows. When a person moves twice, for instance, motifs are either *ABA*, which represent 6.51% of the observed motifs, or *ABC*, which represent only 1.52% of the motifs.

In this case the fact that a person was not the victim of any crimes for a period of a year does not imply that the person is immune to suffering crime (46; 47).

Formally, let $M_i(t)$ be the number of times that person i has moved between time t_0 and t . Let us assume that if a person moves it does not affect the probability of future migrations. Hence, migrations occur independently and we assume that the rate at which the person moves is constant. Therefore, the number of migrations follows a Poisson distribution with rate $\lambda_i t$,

$$M_i(t) \sim Po(\lambda_i t). \quad (1)$$

Both are strong assumptions with respect to migration (independence of observations and constant rate) that are not observed in reality (48). It is known that the longer a person lives in a single location, the lower the probability that the person will leave their city (21; 22). Hence, a migration is more likely to occur after a recent migration (not independence) and a constant rate is not necessarily the case. Yet, regarding migration as a Poisson distribution with a constant rate enables us to analyse the rate at which individuals move (λ_i) rather than the number of migrations directly. This approach enables us to take into account the fact that while a person might not move as a result of a very small rate (or even $\lambda_i = 0$), this could also be the result of a high rate and simply "luck". Hence, instead of counting the number of migrations directly, we observe migration through the lens of "speed". In turn, issues with the arbitrary starting point and observation window are less relevant.

Let $L_i(t)$ be the number of distinct locations that individual i has lived in since time t_0 . Then, if $M_i = 0$ (so, the person has not moved) the number of locations is $L_i = 1$. With $M_i = 1$, then $L_i = 2$, since the person moved between two locations. But, if $M_i = 2$, then on the second migration the person might have moved to their first location (ABA), or might have moved to a new location (ABC). Assume that for each migration after the first one, the person decides whether to move back to a previously known location, with probability π , called the "return rate", or moves to a new location with probability $1 - \pi$. Then, the conditional distribution of $L_i(t)$ given $M_i(t) = m$ is given by

$$L_i(t)|[M_i(t) = m] \sim Bin(m - 1, 1 - \pi) + 1, \quad (2)$$

if $m > 1$, so the person moves more than once, and

$$L_i(t)|[M_i(t) = m] = m + 1 \quad (3)$$

if $m = 0$ or 1 , so the person does not move, or moves only once.

It is easy to show that a Binomial distribution, conditional on a Poisson distribution, also follows a Poisson distribution, with the combined rates and probability π , so that

$$L_i(t) \sim Po(\lambda_i(1 - \pi)t) + 1, \quad (4)$$

if the person moves at least once.

With our method we can estimate the speed (or rate) at which migration occurs λ_i , as well as the rate at which a person moves to a new location, $\lambda_i(1 - \pi)$ (i.e., the migration rate discounted by the return rate $1 - \pi$). Both rates enable us to describe migration patterns in a succinct manner (Figure 2) and to extend the observed patterns outside the limits of the data (either through simulations or through the analytical expressions of the corresponding distributions).

Although

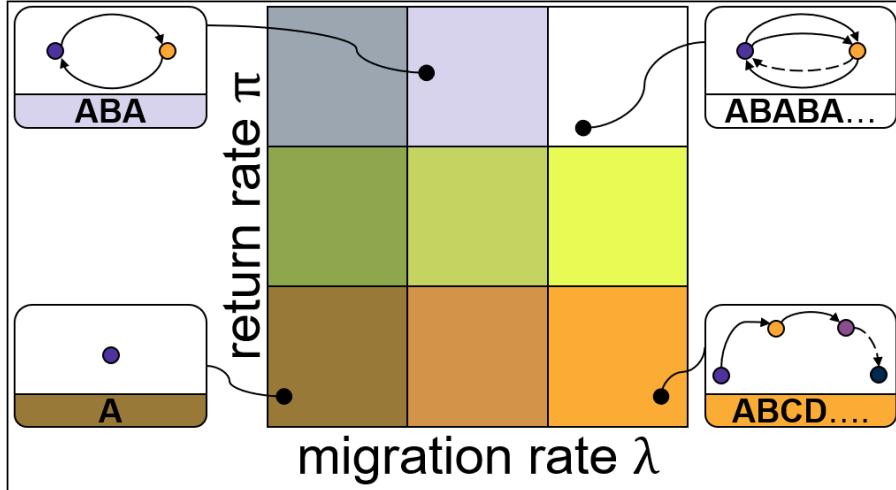


Figure 2: The model has two parameters, λ (horizontal axis) which represents the rate at which a person moves, and the return rate π (vertical axis), which is the propensity that a person returns to a previous city. Different signatures are obtained on distinct regions of the parameter space, where with high migration rate and high return rate, for example, the observed signatures are $ABABCAB\dots$, a medium migration and a medium return rates represents shorter signatures with fewer repetitions, such as $ABCD\dots$, and with a low migration rate, often only A or AB are obtained as signatures.

2.3 Classifying individuals based on their migration rate

The individual migration λ_i and the return π rate enables us to quantify and reproduce migration motifs using a minimal number of parameters. This, in turn, enables us to compare migration rates between distinct populations, such as men and women, the young and the mature. We combine individuals using a mixture model (49), a novel technique in mobility studies, which groups individuals based on their rates. This technique is frequently used in medical studies to divide populations into distinct groups (49; 50), and has also been applied to study the victimisation rates. In this case, instead of migrations the number of crimes suffered by each person, and its associated rate, is studied (47; 46).

Using a mixture model, we group individuals based on their migration rate into k groups, such that $0 \leq \lambda_1 < \lambda_2 < \dots < \lambda_k$ and with relative sizes q_1, q_2, \dots, q_k such that $\sum q_j = 1$. All individuals are assigned to a single group with the same migration rate. The number of groups k is determined by the data, and not known a priori. In this case, distinct groups may correspond to ‘types’ of individuals in terms of migration, i.e., those who move frequently or rarely are likely to be grouped together.

For a population, its corresponding parameters are estimated as follows. We fit a finite mixture model (51; 49; 52) which takes as input the number of migrations of each individual, M_1, M_2, \dots , and gives the number of groups, \hat{k} and the corresponding rate $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k$ and size $\hat{q}_1, \hat{q}_2, \dots, \hat{q}_k$ of each group. Then, the best fit return rate π is estimated from the data by generating a simulated sequence of nine characters (which represents a motif for a person), and minimising the mean square error between the observed and simulated frequencies for a range of π . The return rate is also compared to scenarios with discouraged ($\pi = 0$), preferential ($\pi = 1$) and random returns ($\pi = 1/N$), where N is the number of locations. Results show that the models with no return

migration and with random returns cannot reproduce the observed frequency of motifs (see the Appendix for more details).

Hence, with just a small set of parameters (the migration rate of the group λ_j , its size q_j and the return rate π , which is assumed to be the same for the whole population), we can summarise movements and measure return migration for distinct population groups. Once the parameters are obtained, after a period of τ years, a person expects to move $\tau\lambda_P$ times and expects to live in $\tau\lambda_P(1 - \pi_P) + 1$ distinct cities for values of $\tau > 0$.

2.4 Most individuals will never migrate

Migration is highly concentrated among a few individuals. The mixture model applied individuals in our dataset yields $\hat{k} = 3$, meaning that based on the number of migrations, people can be divided into three groups (relative sizes and rates are shown in Table 2). Our results show that 69.3% of people will not move and can be considered as “stayers”, and that there is a small group containing 2.5% of the individuals who move very frequently, at a rate of $\hat{\lambda}_3 = 3.04$, who can be considered “supermigrants”.

Group	\hat{q}	$\hat{\lambda}$	$\hat{\lambda}(1 - \pi^*)$
1 stayers	0.6926	0.0000	0.0000
2 migrants	0.2826	1.2827	0.3077
3 supermigrants	0.0248	3.0331	0.7275

Table 2: Migration profile of Colombia. The finite mixture model divides the population into three groups: one containing nearly 70% of the population that does not move, and one with less than 2.5% of the population that moves very frequently.

The rate of return migration is $\hat{\pi}^* = 0.7612 \pm 0.0007$ (with intervals obtained through bootstrapping) meaning that after the first migration, roughly three out of four times a person moves back to a previous city. Our results indicate that there is a strong preferential return migration, far from the no-return values with $\pi_H = 0$, and from random returns $\pi_R = 0.016$ but also far from the case with only returns $\pi_O = 1$ (see the Appendix for the details). The distribution of the number of distinct locations of a person in group 3, the supermigrants, for instance, is $L_3 \sim Po(0.7275) + 1$, which means that roughly 18% of the population from that group expects to live in three or more cities.

The top 5% migratory individuals accumulate nearly 25% of the migration rate in Colombia and the top 24% accumulate 80% of the migration rate. Similar to the dichotomy observed for daily mobility patterns (30), here we observe a large population group which will never move and a small population which moves with a high intensity, frequently returning to previous cities.

We can deploy this methodology to uncover migration rates for different population partitions, e.g., men and women; low and high income and young and elderly, along with their return rates π_m and π_w respectively. This enables us to shed light on distinct migration behaviours observed for sub-population groups.

2.5 Women move less and return more frequently

We divide the population by gender and consider each separately. In the data, 61.4% of individuals are male and 38.6% are female. For both genders, the mixture model

gives $\hat{k} = 3$ groups, meaning that both men and women can be subdivided into three groups. The migration profile of each gender (or any subpopulation) is determined by the set of pairs (q_j, λ_j) for each group. These pairs provide an overall description of the migration intensity and can be visually displayed as a step-wise function (see Figure 3 in which individuals are plotted on the horizontal axis, and their corresponding rate on the vertical axis). We observe that both men and women have a sizeable group who is stayer (65.8% in the case of males and 74.5% in the case of females) and in the case of men, a small group (4.2%) are identified as supermigrants. The average migration rate for men is $\lambda^{(m)} = 0.53$, and the average migration rate for women is $\lambda^{(w)} = 0.29$, meaning that, as it has been observed elsewhere (53; 54) and for other types of mobility (34), men move more frequently than women.

Unlike what was encountered for international migration, where it was observed that men are more likely to move and return than women (55), we find that women are more likely to return ($\pi^{(w)} = 0.8161$) than men ($\pi^{(m)} = 0.7407$). Men are more likely, however, to explore a new city. Comparing, for instance, the most migratory groups between men and women, the expected number of distinct locations is 45% larger for men than for women. The motifs per gender (Figure 4) highlight that women are more likely to stay in the same city, to move only once, if they move, to travel slightly shorter geodesic distances (5% shorter on average). If they move twice or more, they tend to return to previous cities.

The migration profile is a multi-dimensional description of migration rates at a population level. However, given two distinct migration profiles (say, men and women) it might not be easy to detect if the rates are more or less concentrated in the population. The concentration of migration is computed as the Gini coefficient of the migration rates from the population (47). Since the Gini coefficient has scale independence (some population groups could be more migratory as a whole) and population independence (it does not matter how large the population is) it is a comparable metric for the concentration of migration. We find that the concentration is higher for women (0.75) than it is for men (0.69). This is because women move at a lower rate than men and those who move do it at a lower rate with a higher chance of returning than men.

2.6 Mature and low-income people move less and return more

Age is a key driver of migration (55), whereby young adults have the highest migration intensity (53) but with significant variation according to age (56). For example, younger adults tend to move to large cities to benefit from density, jobs and amenities, but elderly people frequently move in the opposite direction (57; 58; 59; 60; 53). To investigate these differences in our dataset, as we did with gender, we subdivide the population into two according to their age in 2016. We consider the 25% youngest population (age smaller than 35.9 years) and the 25% oldest population (age larger than 50.9 years) as the young and the mature population respectively (Figure 3). The young population is subdivided by the mixture model into three groups, where 62.9% of the population is a stayer and 1.8% is a supermigrant, with a rate of $\lambda_3^{(y)} = 3.2$. The mature population is subdivided into four groups, where 76.1% of the population is stayer, and 2.2% of the population is supermigrant, with a migration rate $\lambda_4^{(r)} = 3.1$.

Hence, the young population is nearly twice as migratory as the mature population and has a more explorer profile as has been observed for international migration (55). Combined, the young population expects to live in 16% more distinct locations than

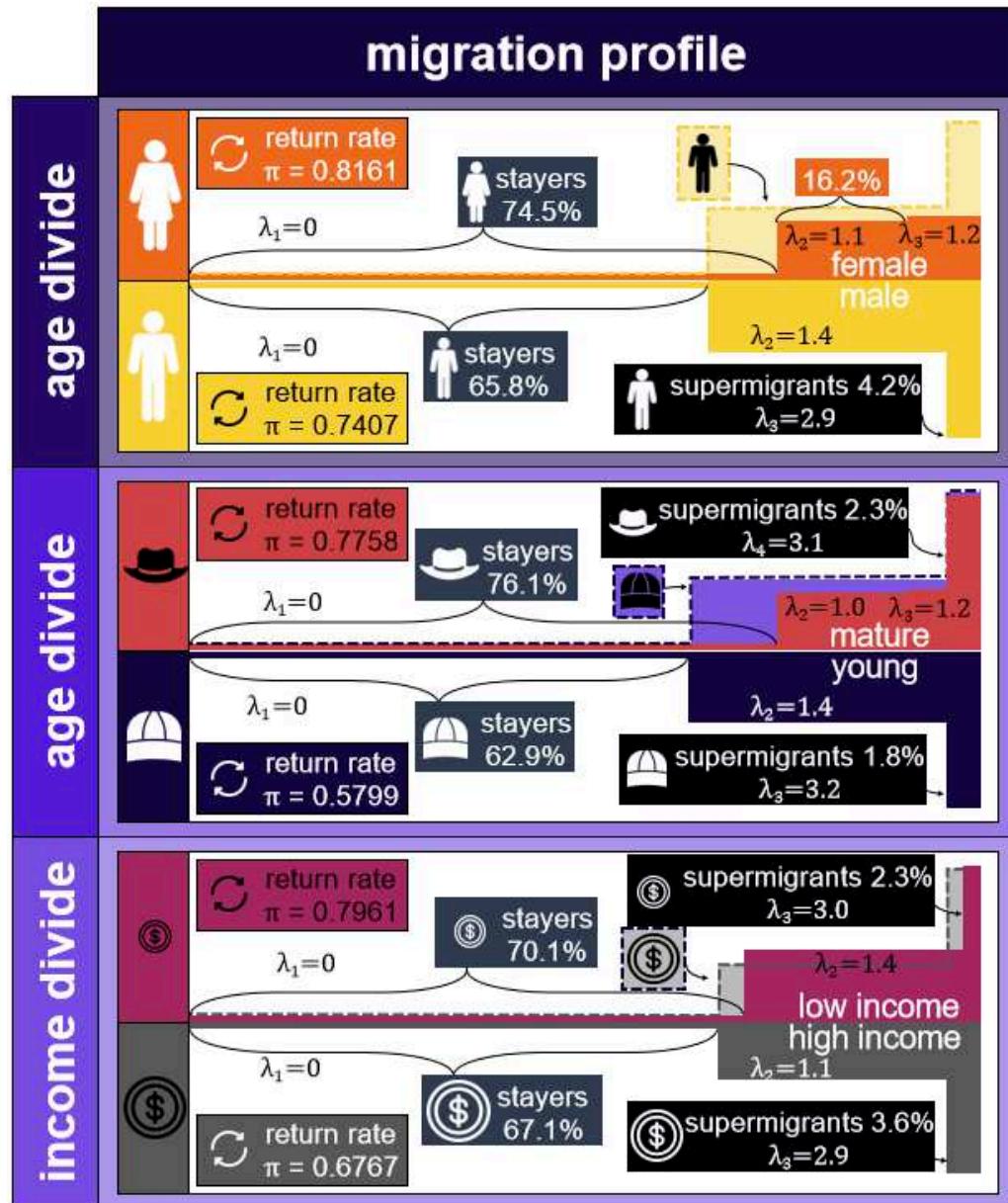


Figure 3: Migration profile for women (upper part) and for men (lower part) in the upper panel, where all individuals are sorted (horizontal axis) according to their migration rate (vertical axis). A similar comparison in the middle panel for young and mature populations. In the bottom panel, the migration profile of the population with lower and higher income.

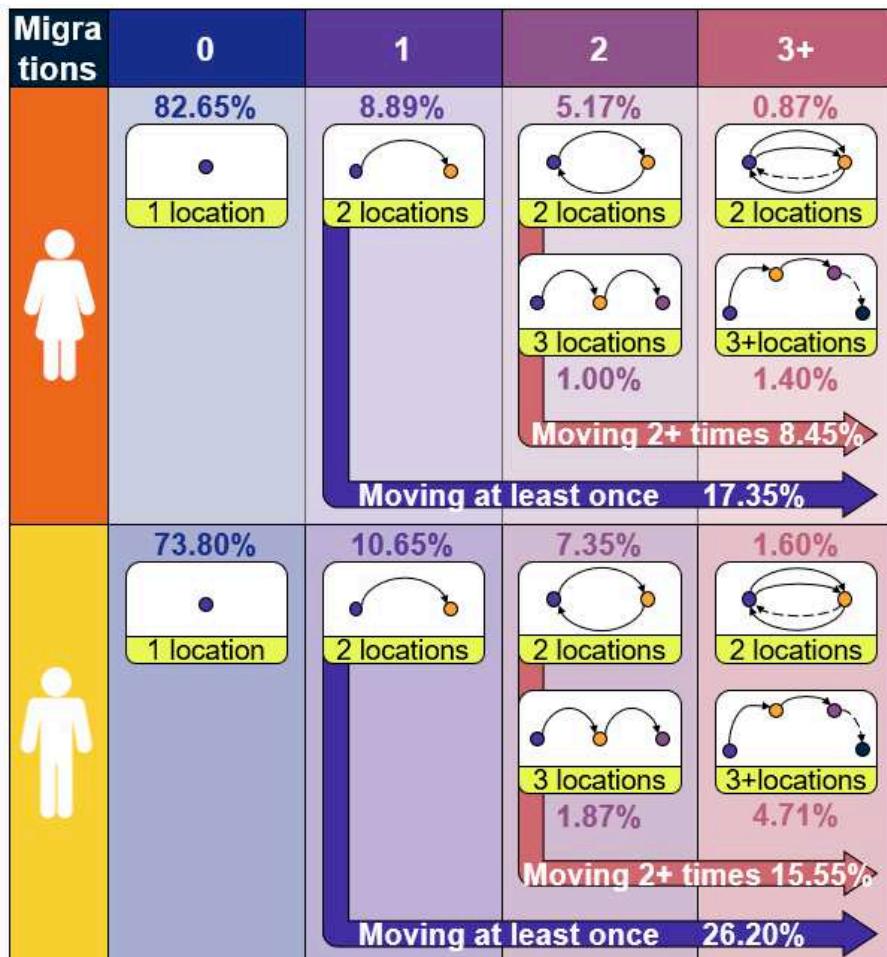


Figure 4: Observed migration motifs for women (top panel) and for men (bottom panel). Women who move twice, for instance, return to their first city (ABA) 84% of the time, and move to a new city (ABC) 16% of the time. Men return to the first city 79% of the time and move to a new city 21% of the time.

the mature population. Unlike what we encountered with respect to the gender divide, the distance between origin and destination for the young and the mature populations is of similar magnitude (223 kilometres each time a person moves).

It is frequently assumed that income or employment are some of the main drivers of migration (61), whereby a person compares their expected income with and without moving (and among distinct destinations) and chooses the best option (62; 63; 64; 2). Migrants are frequently thought to be 'pushed' out of areas with lower income and attracted to areas with higher earnings (65; 66). Still, as it has been noted, expectations might not materialise or people could move for non-monetary reasons and so a person might not earn more income after migration (67). Surprisingly, dividing the population between the 25% with the highest and lowest income in 2016 does not yield such a strong division as much as age or gender (Figure 3). The average migration rates are $\lambda^{(l)} = 0.44$ for low income individuals and $\lambda^{(h)} = 0.42$ for high income individuals and profiles for both groups are similar. A more significant difference is observed in the return rate since the low-income population is more likely to return to previously-lived cities.

Workers in bigger cities are usually more productive and earn more than workers in smaller cities and rural areas (68; 69) although large cities disproportionately attract both high- and low-skilled workers (70). Here, we observe - perhaps counter-intuitively - that (at a country level) people with high and low income have similar migration patterns in terms of their rates and the motifs generated when moving, meaning that a person is almost equally likely to move within a country and to return if they have a low or a high income. Next, we investigate these age and gender divides at a city level, comparing large wealthy cities in the north of Colombia to mid-size and typically poorer cities in the south.

2.7 People from large cities move less and return more frequently

Beyond migration profiles for different population groups (such as women, men, young or mature), motifs also enable us to compare migration profiles between distinct cities. For a city (say A) we compute motifs similarly by looking at the frequency at which a person stays from one year to the next one ($AA\dots$), migrates ($AB\dots$) and returns ($ABA\dots$) and construct the migration profile of the city as follows. The probability of moving μ_A is the ratio between the number of migrations and the number of years in which locations are known. The return rate ρ_A is the frequency at which people who moved from city A return to live in it. Results show that people from large cities, such as Bogotá, Medellín and Cali, have a small probability of moving. Also, if a person moves from these cities, there is a very high chance that the person will return (Figure 5). In smaller cities, the opposite happens. There is a 25% chance that a person from Buga, for example, will move, and only 1 in 8 of those who move will return to that city. As has been seen before (71), there are nonlinear effects between city size and migration, and people from larger cities are much less likely to move and more likely to return than their small-city counterparts. Movements between distinct areas of the countryside are not detected by our method, but we do observe that people from the countryside more generally are very likely to move (to a city) and are 36% less likely to return to any part of the countryside than people from Bogotá.

Next, we apply our methodology to different population groups within the largest cities (bottom panel of Figure 5). The gender and the age divide observed at a national

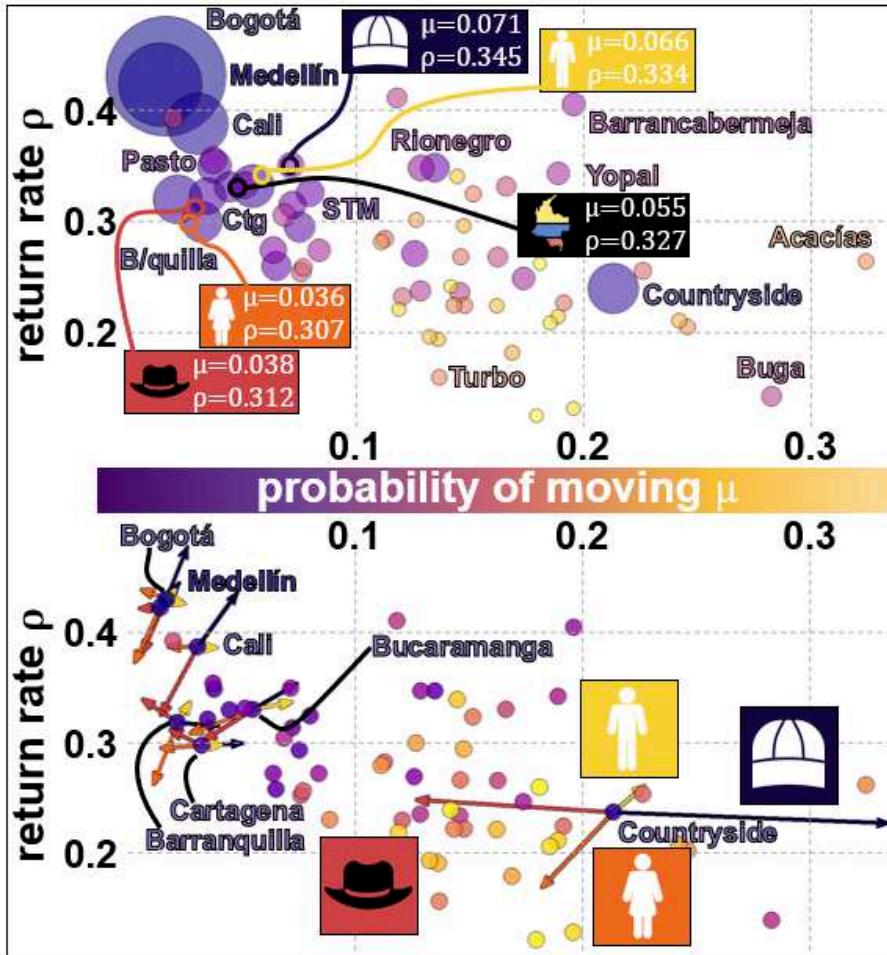


Figure 5: Probability of moving μ (horizontal axis) and return rate ρ (vertical axis) for the 62 metropolitan areas and the countryside in Colombia (top panel). The bottom panel shows the same points, but in this case our analysis considers four population groups (men, women, young and mature) for a subset of the largest cities separately. The coloured arrows indicate the probability of moving μ and return rate ρ for men (yellow), women (orange), the young (blue) and the mature (red) for each of these cities.

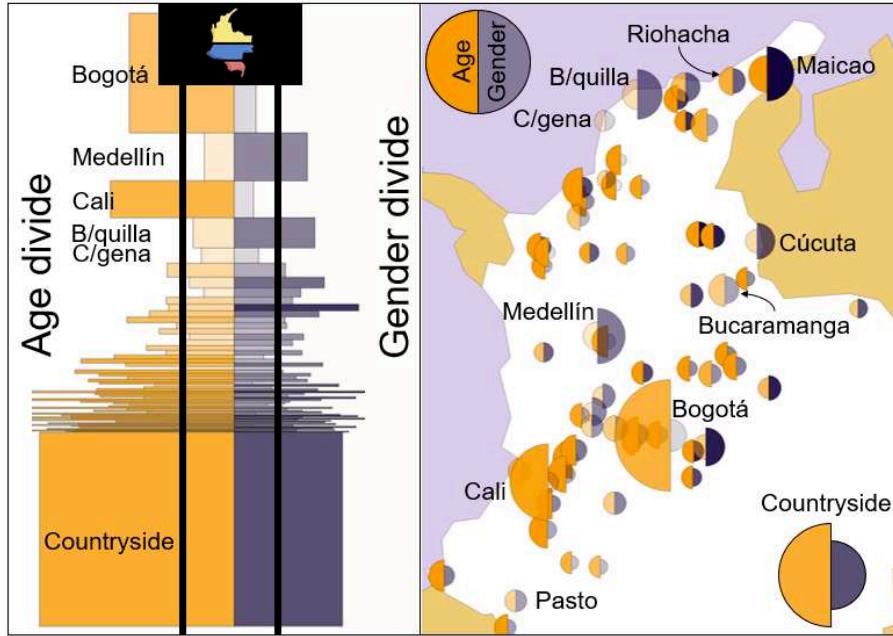


Figure 6: Age divide per city where the height of each bar (left) or the size of the disc (right) represents the city size. Longer bars correspond to a higher divide, and the national level of age and gender divide is represented by the vertical dark lines.

level are also observed at a similar scale at a city level. Overall, we find that women and mature populations tend to move less, while young and male groups tend to move more. A young person from the countryside, for example, is 2.4 times more likely to move than a mature person, and 7% less likely to return.

Thus far, we have observed significant age and gender differences in migration motifs. Yet, this marked difference is not homogeneously observed across the country. In some cities, men and women have similar migration and return rates, whereas in other parts, particularly small cities and countryside, the gender - and age - divide is more pronounced. This type of city-level disparity is the mechanism through which accelerated ageing and gender imbalance in a city is deepened in some parts of the country (11; 9). We measure the age divide for each city as the Euclidean distance between the probability of moving μ and the return rate ρ between the mature and the young population groups. The gender divide is measured similarly as the distance between the probability of moving μ and the return rate ρ between women and men. A larger divide, whether it is age or gender, means more distinct migration patterns between the population groups.

We find that the age and gender divide are not homogeneous across Colombia (Figure 6). In Bogotá and Cali there is a large age divide, indicating that young people from those two cities move more frequently, but both cities have a smaller gender divide than the national average. The opposite happens in Medellín and Barranquilla, where the gender divide is much larger than the age divide. In fact, Medellín has the smallest age divide. In general terms cities from South Colombia have a greater age divide than a gender divide, whereas in North Colombia, age divide and gender divide are of similar magnitude. In the North part of Colombia many large and industrialised cities attract nearby workers, whereas in the rural parts of the South the age and gender divide is much more substantial than most parts of the country.

3 Discussion

Internal migration is critical for urban dynamics. With the progressive convergence of birth and death rates between countries, migration is the principal source of population re-distribution within countries (7). Migration is a specific type of mobility with a lower frequency and much more extended periods, which in turn, poses a severe challenge to most data-driven models, as individuals need to be observed for many years to detect aspects such as the concentration of migration and return patterns. Access to an administrative dataset enabled us to consider nine years for individuals and their internal migration in Colombia. Far from being a representative sample of the country, it enabled us to understand internal migration for a specific set of people: formal employees who worked (almost) continuously between 2008 and 2016. It is a biased sample of the population in terms of age, income, gender and other demographics, but we have a complete set of locations for each person. We designed a generative algorithm for reproducing migration motifs based on population-level parameters of onwards and return migration.

A frequent challenge when analysing migration is that data is usually available for a restricted time window with an arbitrary starting point (2008 in our case) and so an individual might not have moved in the observed time window but could have recently moved or might move soon after. Furthermore, the first observed movements might not be identified as return migrations since we might not know the first locations. Our method allows backcasting and forecasting data and migration patterns, as it is based on estimated speeds (of migrating and returning) so that observations can be constructed outside the observed time limits of the available data. A similar methodology can be applied to analyse job switching or the evolution of a criminal career, whereby the observed motifs between different types of crime could reveal the progress between minor offences and violent crimes.

The concentration of migration and return migration are two key ingredients to capture internal migration patterns. Results show that 77% of the individuals observed did not move for nine years, so we consider them stayers, but some moved many times, so migration is highly concentrated among a few individuals. The concentration of migration is high, with a coefficient of 0.7206 for all Colombia, but it is even more concentrated among the mature (0.7942) and female (0.7501) populations. Return migration is widespread. Three out of four times a person moves, after their first migration, they will move back to previously lived cities. Thus, a large part of the internal migration flow is going back. Still there is a large difference between young and mature, between men and women and between cities. Young people from the countryside are highly likely to move and not going back, as opposed to mature people from large cities, who are likely not to move and if they do, they almost certainly return.

Our results highlight a very substantial challenge in terms of gender. We observe that women suffer from a wage gap (72), are less likely to belong to be formal employees (40% of the formal market in 2016 were women), and are also less likely to pass our filtering procedure (38.6% of the filtered observations), which indicates that some women are slightly more likely than men to work only for a few years in the formal economy or to work intermittently. Women are more likely to stay in the same city; women who move, do it at a lower rate than their male counterparts, have a higher tendency of returning to previously-lived cities, and even in terms of geodesic distance, female migration is slightly shorter. The gender divide exhibits a larger difference than the age

divide. Most of these differences might be rooted in the country's gender inequality, with substantial implications for Colombia's mobility and productivity.

Finally, we also observe a challenge for secondary cities in Colombia in terms of their migration patterns. Smaller cities and the countryside tend to combine a larger outflow of people with fewer returns, compared to Medellín, Cali and Bogotá. Secondary cities in Colombia create less formal jobs and have a smaller formality rate (73), and in our data, we are only observing formal employees who remain in the labour market for consecutive years. Thus, small secondary cities are lagging metropolitan areas, less capable of creating formal employment, and are less capable of keeping their workers or attracting them to return after their first migration and compete with primary cities (74). The challenge is substantial, as 68.3% of the population from Colombia lives in one of the 62 metropolitan areas considered here, with 43% residing in a city with more than one million inhabitants and 57% in smaller cities. Most Colombians live in a secondary city with less than one million inhabitants (25.2%) or the countryside (31.2%), where people are more inclined to leave and never return. Our results indicate that small cities in Colombia will suffer an accelerated ageing process due to internal migration. A similar process is also expected in other parts of the world, as countries experience a higher intensity of inter-urban migration (75) and so it is likely that small cities from Mexico or the rural area from Brazil, for example, experience a similar migration process and thus, the same accelerated ageing problem as a result.

4 Methods

4.1 Defining metropolitan areas

Colombia is divided into 1,122 municipalities, some with a population of a few million and the three smallest municipalities with less than 1,000 inhabitants in 2018. Some of the municipalities are part of the same metropolitan area, such as Medellín and Envigado, and many of the municipalities are rural. We merge urban municipalities into metropolitan areas, spatially delineated through commuting patterns, according to previous research (76; 73).

The metropolitan area of Bogotá, for instance, consists of 23 municipalities, including the municipality of Bogotá itself, Soacha, Facatativá and others. In total, 19 metropolitan areas are formed as the union of two or more municipalities (the four largest metropolitan areas are Bogotá with 9.7 million inhabitants; Medellín with 3.9 million; Cali with 2.9 million; and Barranquilla with 2.4 million inhabitants) and are labelled with the name of the largest municipality. Movements of people within the same metropolitan areas are ignored (for example, if a person moved from Soacha to Bogotá it is not considered migration).

Also, 43 municipalities which are not part of the 19 metropolitan areas but are urban are considered as separate "cities", including for example Ibagué and Santa Marta (with more than 500 thousand inhabitants). In total, we obtain 62 "cities" with this method (19 metropolitan areas and 43 urban municipalities), and use the term cities to refer this set.

Municipalities which are not added by this method are considered to be "rural". In total, 964 municipalities are labelled as rural and movements between different parts of the rural areas are not considered, although movements between the countryside and cities are studied.

Observation	Kept?	Imputation	Motif
○○A AAA AAA	No	—	—
AAA ○○B BBB	No	—	—
AAA AAA A○○	No	—	—
○AA A○A A○A	Yes	AAA AAA AAA	A
AAA ○BB BBB	Yes	AAA ABB BBB	AB
○A○ B○C ○Do	Yes	AAA BBC CDD	ABCD
AB○ CD○ A○E	Yes	ABB CDD AAE	ABCDAE

Table 3: Distinct scenarios on the filtering and the imputation based on the observed data for each year.

In total, migration between 63 distinct locations is studied: 19 metropolitan areas; 43 urban municipalities and the countryside.

4.2 Data selection and missing information

There are 16,576,254 observations in the PILA dataset. Any formal employee who worked for at least one month on any given year between 2008 and 2016 appears on the dataset. However, to detect migration patterns, we need to know their location across a number of years. Since some of the individuals appear only a few years, we need to filter out individuals for whom not enough information is known.

For each year, the most frequent location of each person is selected. For some individuals, their location during nine years is known (so that we have full information), but for some, there is some missing data, which includes people who retired between 2008 and 2016, a person who did not work for a period of time or who joined the labour market after 2008.

It is impossible to detect migration for individuals in which their location is known only for a few years. Observations are filtered and missing information imputed trying to keep as many observations as possible, but avoiding biases in the number of locations and migrations. The procedure is as follows:

- Observations are dropped if they have any two consecutive years of missing information.
- For the remaining individuals, if there is one missing year, the missing location for that specific year is imputed by the location of the previous year.
- If an individual is missing the location of the first year, it is imputed by the location of the second year.

Schematically, if we represent with ○ the missing year, we have the following filtering and imputation scenarios (Table 3).

Notice that with this filtering and imputing process, individuals who could have lived in a city for more than 12 months without it being detected in our dataset are dropped (as the first three examples on the table, which could be an undetected migration) but we keep individuals for whom undetected migration is not possible. The imputing procedure does not increase the number of migrations and does not alter motifs. For instance, on fourth row of Table 3, there are three missing years and they are all imputed

with A which assumes that the person did not move. In the fifth row, for which there is one missing year, this alternatively be imputed with A or B but it would not change the number of migrations or locations of the individual, or the migration motif. An individual is kept if they are missing up to five years but those are not consecutive (as the sixth row), and the procedure also keeps return migrations (as the seventh row of the table).

4.3 Reproducing complex motifs

Internal migration patterns are complex and they are the emergent result of millions of people deciding to move for individual reasons, but exhibiting collective behaviour. A generative algorithm is constructed for the sequence of cities of each individual, which takes as input the migration motif of each person and a model parameter called the return rate, which is estimated as follows. Firstly, individuals are grouped based on a mixture model into an unknown number of groups, each group of individuals with the same migration rate λ_j . A sampled probability $\pi \in [0, 1]$ is used to simulate motifs. For person i , with migration rate λ_i , the number of migrations in τ years is sampled from $M_i \sim Po(\tau\lambda_i)$. After the first migration, and each time a person moves, they move back to any of the cities in which they previously lived with probability π . If there is more than one city in which they previously lived, it is randomly chosen. The algorithm generates a sequence of M characters, which is then transformed into a (simulated) motif. For a population of P individuals, we compare the observed frequency of observed and simulated motifs and keep the value of π which produces the best fit to the data ($\hat{\pi}$ i.e., the value which best reproduces the observed frequency of motifs).

We also compare the frequency of each motif with two distinct models. A null model, with no return migration, so $\pi_H = 0$, and so each time the person moves, they choose a different city with no return migration, so motifs arise such as A , AB , ABC and so on. We also compare with a second model with complete randomness in the sequence of cities, so each time the person moves, they select any of the $N - 1$ cities randomly and moves. Under complete randomness, after the first migration, return migration could occur if the person randomly chooses a city in which they previously lived. If $\hat{\pi}$ gives values close to zero, we observe discouraged returns, meaning that individuals rarely move back. In the extreme case, if $\hat{\pi} = 0$, we obtain the model with no return. With values of $\hat{\pi} \approx 1/(N - 1)$ we observe random returns, meaning that return migration happens at a similar frequency in which randomness would happen. With high values of $\hat{\pi}$ we observe preferential returns, meaning that people return more frequently than randomness to previous cities, where the extreme case of $\hat{\pi} = 1$ produces only motifs with two cities, with the form $ABAB\dots$.

After estimating the best fit of $\hat{\pi}$, we observe that the model with a return rate $\pi_H = 0$ is capable of dividing those who move and those who do not move, and produces some of the onward migration patterns (AB , ABC and so on) although it tends to overestimate their frequency (Figure 7). The return model (with a return rate $\pi_R = 1/(N - 1)$) captures some of the frequency of return migration, but it overestimates the motif ABC and underestimates the motif ABA , as it does not favour return migration. This behaviour is best captured by our model, named the Poisson return model, which mimics quite well the motifs generated by migrants in Colombia. For the frequently observed motifs, the Poisson return model does capture most of the frequency, so the model is capable of reproducing the frequency of complex motifs,

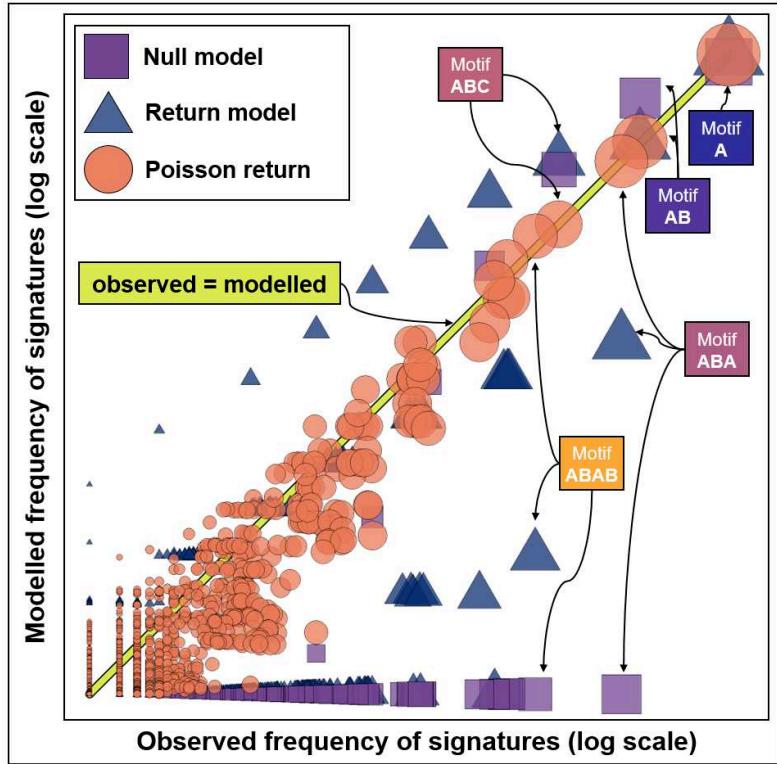


Figure 7: The horizontal axis is the observed frequency of motifs (in a logarithmic scale) and the vertical axis is the modelled frequency. The yellow line is the identity, and observations closer to the line have a better fit. The size of each mark is proportional to the frequency, so that more frequent motifs have a bigger mark.

such as $ABCADA$ or others.

The most frequently observed motifs are captured with the Poisson return model with a value of $\hat{\pi} = 0.7612$, with a return rate much higher than the random model $\pi_R = 0.0161$, meaning that return migration happens much more frequently than under a model assuming random movements. Return migration is indeed much more frequent than a null hypothesis and people are much more likely to go back than move to new locations. The Poisson return model is also a generative algorithm since it is possible to simulate migrations for longer (or shorter) periods of time.

4.4 Return rate for different subgroups

The rate of return migration for the whole population is $\pi^* = 0.7612$, obtained by minimising the mean square error between the observed and the modelled motifs using the Poisson return model (Figure 8). For other subgroups the return rate, estimated with the same generative algorithm, has different values. The smallest return rate is obtained for the 25% youngest population with $\pi_Y = 0.5799$ and the highest is obtained for women $\pi_W = 0.8161$, so that women are 40% more likely to return than a person from the 25% youngest group.

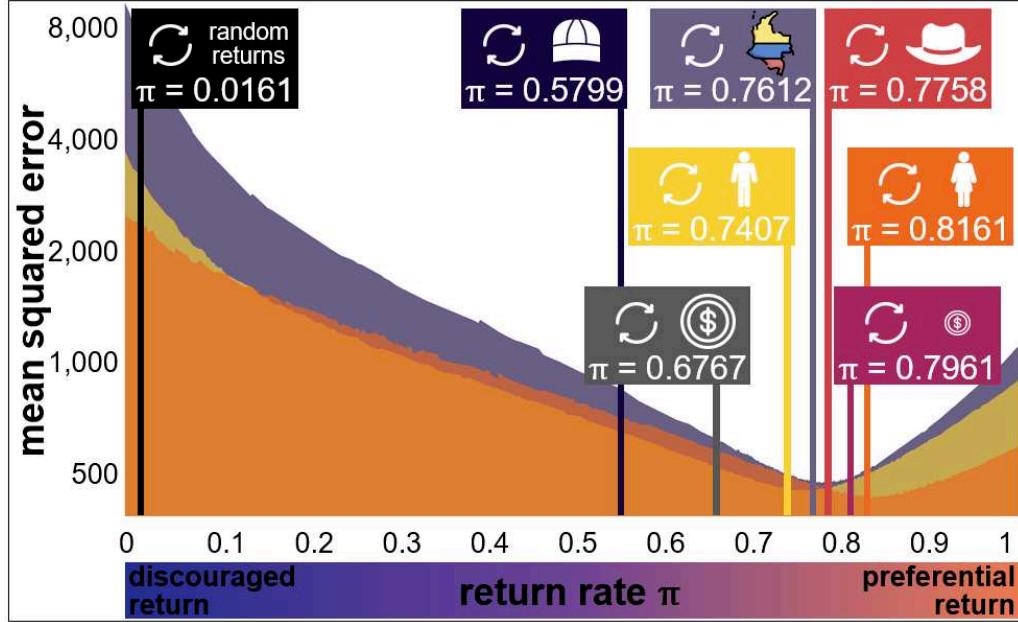


Figure 8: Square error of the motif frequency (observed vs simulated) for different values of the return rate π , where the value $\pi^* = 0.7612$ minimises the error. The same analysis (and corresponding curves) are computed for the return rate for male and female, and similar (curves not shown) for young and mature populations. Random returns would be when $\pi = 1/62 = 0.0161$, so there is a much higher preferential return, particularly for women.

5 Data availability

We use administrative records of the social security system in Colombia (abbreviated as PILA in Spanish, meaning the Integrated Report of Social Security Contributions), which contains job information about all formal workers in Colombia. We are unable to directly share the raw data. However, there is a protocol for gaining secure access to the data via the Ministry of Finance and Public Credit (MFPC) of Colombia or the Ministry of Health and Social Protection (MHSP) of Colombia. We followed that protocol and gained permission to use the PILA dataset for the current study. Please contact the ministries directly for more information.

6 Code availability

No relevant code was produced for this manuscript.

References

- [1] Moritz UG Kraemer, Adam Sadilek, Qian Zhang, Nahema A Marchal, Gaurav Tuli, Emily L Cohn, Yulin Hswen, T Alex Perkins, David L Smith, Robert C Reiner, et al. Mapping global variation in human mobility. *Nature Human Behaviour*, 4(8):800–810, 2020.

- [2] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.
- [3] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.
- [4] Alina Sîrbu, Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Marco Conti, Fosca Giannotti, Riccardo Guidotti, Simone Bertoli, Jisu Kim, Cristina Ioana Muntean, et al. Human migration: the big data perspective. *International Journal of Data Science and Analytics*, pages 1–20, 2020.
- [5] Moshe Levy. Scale-free human migration and the geography of social networks. *Physica A: Statistical Mechanics and its Applications*, 389(21):4913–4917, 2010.
- [6] Vincent Verbavatz and Marc Barthelemy. The growth equation of cities. *Nature*, 587(7834):397–401, 2020.
- [7] Martin Bell, Elin Charles-Edwards, Philipp Ueffing, John Stillwell, Marek Kupiszewski, and Dorota Kupiszewska. Internal migration and development: Comparing migration intensities around the world. *Population and Development Review*, 41(1):33–58, 2015.
- [8] June JH Lee, Jill Helke, and Frank Laczko. *World Migration Report 2015*. International Organization for Migration, Geneva, Switzerland, 2015.
- [9] Steffen Kröhnert and Sebastian Vollmer. Gender-specific migration from Eastern to Western Germany: Where have all the young women gone? *International Migration*, 50(5):95–112, 2012.
- [10] Hannah Ritchie and Max Roser. Gender ratio. *Our World in Data*, 2019.
- [11] Amina Maharjan, Siegfried Bauer, and Beatrice Knerr. Do rural women who stay behind benefit from male out-migration? a case study in the hills of Nepal. *Gender, Technology and Development*, 16(1):95–123, 2012.
- [12] Jorge Rodríguez-Vignoli and Francisco Rowe. How is internal migration reshaping metropolitan populations in Latin America? a new method and new evidence. *Population Studies*, 72(2):253–273, 2018.
- [13] Jorge De la Roca. Selection in initial and return migration: Evidence from moves across Spanish cities. *Journal of Urban Economics*, 100:33–53, 2017.
- [14] Adam Dennett. Understanding internal migration in Britain at the start of the 21st century. University of Leeds, 2010.
- [15] Aude Bernard, Francisco Rowe, Martin Bell, Philipp Ueffing, and Elin Charles-Edwards. Comparing internal migration across the countries of Latin America: A multidimensional approach. *PloS one*, 12(3):e0173895, 2017.

- [16] Yuri Mansury and J.K. Shin. Size, connectivity, and tipping in spatial networks: Theory and empirics. *Computers, Environment and Urban Systems*, 54:428–437, 2015.
- [17] Donata Gnisci. West african mobility and migration policies of OECD countries. Paris, Organisation for Economic Cooperation and Development, Sahel and West Africa Club, 2008.
- [18] WFP. At the root of exodus: Food security, conflict and international migration. World Food Programme, 5 2017.
- [19] Wim Naude. Conflict, disasters and no jobs: Reasons for international migration from Sub-Saharan Africa. Number 85 in 1. UNU-WIDER, 2008.
- [20] Amelie F. Constant and Klaus F. Zimmermann. The dynamics of repeat migration: A markov chain analysis. *International Migration Review*, 46(2):362–388, 2012.
- [21] Robert McGinnis. A stochastic model of social mobility. *American Sociological Review*, pages 712–722, 1968.
- [22] George C Myers, Robert McGinnis, and George Masnick. The duration of residence approach to a dynamic stochastic model of internal migration: a test of the axiom of cumulative inertia. *Eugenics Quarterly*, 14(2):121–126, 1967.
- [23] Ernest George Ravenstein. The laws of migration. *Journal of the Statistical Society of London*, 48(2):167–235, 1885.
- [24] Joshua J Lewer and Hendrik Van den Berg. A gravity model of immigration. *Economics Letters*, 99(1):164–167, 2008.
- [25] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PloS One*, 7(5):e37027, 2012.
- [26] Riccardo Di Clemente, Miguel Luengo-Oroz, Matias Travizano, Sharon Xu, Bapu Vaitla, and Marta C González. Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nature Communications*, 9, 2018.
- [27] Marta C González, Cesar A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- [28] Renaud Lambiotte, Vincent D Blondel, Cristobald De Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul Van Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325, 2008.
- [29] Shan Jiang, Joseph Ferreira, and Marta C Gonzalez. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data*, 3(2):208–219, 2017.
- [30] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. Returners and explorers dichotomy in human mobility. *Nature Communications*, 6:8166 EP –, 09 2015.

- [31] S. Rinzivillo, L. Gabrielli, M. Nanni, L. Pappalardo, D. Pedreschi, and F. Giannotti. The purpose of motion: Learning activities from individual mobility networks. In 2014 International Conference on Data Science and Advanced Analytics (DSAA), pages 312–318, Oct 2014.
- [32] Luca Pappalardo, Maarten Vanhoof, Lorenzo Gabrielli, Zbigniew Smoreda, Dino Pedreschi, and Fosca Giannotti. An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics*, 2(1):75–92, Dec 2016.
- [33] Mariana Macedo, Laura Lotero, Alessio Cardillo, Hugo Barbosa, and Ronaldo Menezes. Gender patterns of human mobility in Colombia: Reexamining Ravenstein’s laws of migration. In *Complex Networks XI*, pages 269–281. Springer, 2020.
- [34] Laetitia Gauvin, Michele Tizzoni, Simone Piaggesi, Andrew Young, Natalia Adler, Stefaan Verhulst, Leo Ferres, and Ciro Cattuto. Gender gaps in urban mobility. *Humanities and Social Sciences Communications*, 11(1), 2020.
- [35] Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.
- [36] John Stillwell, Peter Boden, and Adam Dennett. Monitoring who moves where: Information systems for internal and international migration. In *Population Dynamics and Projection Methods*, pages 115–140. Springer, 2011.
- [37] Julie DaVanzo. Repeat migration in the United States: who moves back and who moves on? JStor, 1980.
- [38] Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.
- [39] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
- [40] Xu Hong-Lin, Yan Han-Bing, Gao Cui-Fang, and Zhu Ping. Social network analysis based on network motifs. *Journal of Applied Mathematics*, 2014, 2014.
- [41] Toby Davies and Elio Marchione. Event networks and the identification of crime pattern motifs. *PloS One*, 10(11):e0143638, 2015.
- [42] United Nations High Commissioner for Refugees UNHCR PROGRAMME. Global report, 2019, 2019.
- [43] Allen C Kelley and Leonard W Weiss. Markov processes and economic analysis: the case of migration. *Econometrica: Journal of Econometric Society*, pages 280–297, 1969.
- [44] Neil W Henry, Robert McGinnis, and Heinrich W Tegtmeyer. A finite model of mobility. *The Journal of Mathematical Sociology*, 1(1):107–118, 1971.
- [45] Marc Barthélemy. *Spatial networks*. Springer, USA, 2014.

- [46] Rafael Prieto Curiel, Sofía Collignon Delmar, and Steven Richard Bishop. Measuring the distribution of crime and its concentration. *Journal of Quantitative Criminology*, pages 1–29, 2017.
- [47] Rafael Prieto Curiel and Steven Richard Bishop. A measure of the concentration of rare events. *Scientific Reports*, 6, 2016.
- [48] David A Plane, Christopher J Henrie, and Marc J Perry. Migration up and down the urban hierarchy and across the life course. *Proceedings of the National Academy of Sciences*, 102(43):15313–15318, 2005.
- [49] Geoffrey McLachlan and David Peel. Finite mixture models. John Wiley & Sons, 2004.
- [50] Dankmar Böhning, Ekkehart Dietz, and Peter Schlattmann. Recent developments in computer-assisted analysis of mixtures. *Biometrics*, pages 525–536, 1998.
- [51] Peter Schlattmann, Johannes Hoehne, and Maryna Verba. CAMAN: Finite Mixture Models and Meta-Analysis Tools - Based on C.A.MAN, 2016. R package version 0.74.
- [52] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [53] Andrei Rogers and Luis J Castro. Model migration schedules. International Institute for Applied Systems Analysis (IIASA), 1981.
- [54] Adam Dennett and John Stillwell. Internal migration in Britain, 2000–01, examined through an area classification framework. *Population, Space and Place*, 16(6):517–538, 2010.
- [55] Douglas S Massey and Rene M Zenteno. The dynamics of mass migration. *Proceedings of the National Academy of Sciences*, 96(9):5328–5335, 1999.
- [56] Adam Dennett and John Stillwell. Internal migration patterns by age and sex at the start of the 21st century. In *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*, pages 153–174. IGI Global, 2010.
- [57] David A Plane and Jason R Jurjevich. Ties that no longer bind? the patterns and repercussions of age-articulated migration. *The Professional Geographer*, 61(1):4–20, 2009.
- [58] Andrei Rogers. Age patterns of elderly migration: an international comparison. *Demography*, 25(3):355–370, 1988.
- [59] David A Plane. Demographic influences on migration. *Regional Studies*, 27(4):375–383, 1993.
- [60] Aba Schwartz. Interpreting the effect of distance on migration. *Journal of Political Economy*, 81(5):1153–1169, 1973.
- [61] Michael P Todaro. A model of labor migration and urban unemployment in less developed countries. *The American Economic Review*, 59(1):138–148, 1969.

- [62] Guido Dorigo and Waldo R Tobler. Push-pull migration laws. *Annals of the Association of American Geographers*, 73(1):1–17, 1983.
- [63] Andres J Garcia, Deepa K Pindolia, Kenneth K Lopiano, and Andrew J Tatem. Modeling internal migration flows in Sub-Saharan Africa using census microdata. *Migration Studies*, 3(1):89–110, 2015.
- [64] Samuel A Stouffer. Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5(6):845–867, 1940.
- [65] Everett S Lee. A theory of migration. *Demography*, 3(1):47–57, 1966.
- [66] Xingna Nina Zhang, Wenfei Winnie Wang, Richard Harris, and George Leckie. Analysing inter-provincial urban migration flows in China: A new multilevel gravity model approach. *Migration Studies*, 07 2018.
- [67] Larry A Sjaastad. The costs and returns of human migration. *Journal of Political Economy*, 70(5, Part 2):80–93, 1962.
- [68] PP Combes, G Duranton, L Gobillon, and S Roux. Estimating agglomeration effects with history, geology, and worker fixed effects. *The Economics of Agglomeration*, Nationla Bureau of Economic Research, Cambridge, MA, 2010.
- [69] Jorge De La Roca and Diego Puga. Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142, 2017.
- [70] Jan Eeckhout, Roberto Pinheiro, and Kurt Schmidheiny. Spatial sorting. *Journal of Political Economy*, 122(3):554–620, 2014.
- [71] Rafael Prieto Curiel, Luca Pappalardo, Lorenzo Gabrielli, and Steven Richard Bishop. Gravity and scaling laws of city to city migration. *PloS one*, 13(7):e0199892, 2018.
- [72] Alejandro Hoyos, Hugo Ñopo, and Ximena Peña. The persistent gender earnings gap in Colombia, 1994-2006. *Documento CEDE*, 1(2010-16), 2010.
- [73] Neave O’Clery, Rafael Prieto Curiel, and Eduardo Lora. Commuting times and the mobilisation of skills in emergent cities. *Applied Network Science*, 4(1):118, 2019.
- [74] Brian Herbert Roberts and Rene Peter Hohmann. The system of secondary cities: The neglected drivers of urbanising economies. *CIVIS Sharing Knowledge and Learning from Cities*, 7, 2014.
- [75] Wilbur Zelinsky. The hypothesis of the mobility transition. *Geographical Review*, pages 219–249, 1971.
- [76] Gilles Duranton. Delineating metropolitan areas: Measuring spatial labour market networks through commuting patterns. In *The Economics of Interfirm Networks*, pages 107–133. Springer, 2015.

7 Acknowledgements

This project received financial support from the PEAK Urban programme, funded by UKRI's Global Challenge Research Fund, grant no. ref.: PES/P011055/1.

8 Author contributions

R.P.C. designed the study. M.Q.D. processed the data. E.L. and N.O. analysed the results. All authors wrote the manuscript.

9 Competing interests

The authors declare no competing interests.

10 Materials and Correspondence

All correspondence should be addressed to rafael.prieto.13@ucl.ac.uk

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MigrationmodelsApp.pdf](#)