Journal of Urban Economics 119 (2020) 103277

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

# Calling from the outside: The role of networks in residential mobility☆

Konstantin Büchel[a,1], Maximilian V. Ehrlich[a,2], Diego Puga[b,3], Elisabet Viladecans-Marsal[c,4]

[a] *Department of Economics and Center for Regional Economic Development, University of Bern, Schanzeneckstrasse 1, Bern 3001, Switzerland*
[b] *CEMFI, Casado del Alisal 5, Madrid 28014, Spain*
[c] *Department of Economics, Universitat de Barcelona and IEB, 1-11 John M Keynes, Barcelona 08034, Spain*

ARTICLE INFO

*JEL classification:*
R23
L14

*Keywords:*
Residential mobility
Social networks

ABSTRACT

Using anonymised cellphone data, we study how social networks shape residential mobility decisions. Individuals with few local contacts are more likely to change residence. Movers strongly prefer neighbourhoods where they already know more people nearby. Contacts matter because proximity to them is valuable and makes attractive locations more enjoyable. They also provide hard-to-find local information and reduce frictions, especially in home-search. Effects are not driven by similar people being more likely to be friends and move between certain locations. Recently-moved and more central contacts are particularly influential. With age, proximity to family gains importance over friends.

## 1. Introduction

This paper uses cellphone Call Detail Records to study the role played by the location of a person's social network in determining whether to change residence and to which city and neighbourhood. The decision of where to live is of fundamental economic importance. We spend about two-thirds of our time at home and around one-third of our income buying or renting that home. Depending on our residential location choice, there are also substantial differences in with whom we can interact as well as in the extent to which jobs, education opportunities and amenities are within reach. Even when accessible, getting to people and places often requires substantial transit, and we typically spend close to 10% of our wake time travelling, with considerable variation around this figure according to where we live. As circumstances change, so do our residential location choices, and in many countries, 5% or more of the population moves each year.[5]

Research on residential location choices tends to focus on determinants that are common across individuals or broad groups, such as job opportunities, housing costs, amenities, accessibility, and taxes. These common determinants create benefits and costs that tend to balance out across locations. When shocks alter this balance, individuals react by relocating from worsened to improved locations (Blanchard and Katz, 1992). Relocation flows then change house prices and earnings until a spatial equilibrium is restored (Rosen, 1979; Roback, 1982; Glaeser, 2008).

☆ We are very grateful to our data provider for providing the facilities and data to conduct this research project; we are particularly indebted to Imad Aad and Holger Müller, who accompanied the project. We also thank Matthew Turner for a very helpful discussion and search.ch, guidle.com and Meta-Sys, for providing data on travel times, local events, and rental prices. Puga acknowledges funding from the European Research Council under the European Union's Horizon 2020 Programme (ERC Advanced Grant agreement 695107 – DYNURBAN). Viladecans-Marsal acknowledges funding from Spain's Ministry of Science and Innovation (grant ECO2016–75941–R) and the Catalan Research Agency (grant 2017SGR796).

*E-mail addresses:* konstantin.buechel@vwi.unibe.ch (K. Büchel), maximilian.vonehrlich@vwi.unibe.ch (M.V. Ehrlich), diego.puga@cemfi.es (D. Puga), eviladecans@ub.edu (E. Viladecans-Marsal).
[1] website: http://www.vwi.unibe.ch/about_us/people/dr_buechel_konstantin/
[2] website: http://staff.vwi.unibe.ch/vonehrlich/.
[3] website: http://diegopuga.org.
[4] website: https://elisabetviladecansmarsal.com.

[5] The average person spends at home 15.6 h per day in the United States, 15.8 h in Canada, and 15.7 h in Germany (Klepeis et al., 2001; Brasche and Bischof, 2005; Matz et al., 2014). According to consumer expenditure surveys, housing accounts for 33% of consumer expenditure in the United States, 29% in Canada, and 27.5% in Switzerland. The average person in the United States spends 80 min per day travelling and 15.2 h awake according to, respectively, the National Household Travel Survey and the American Time Use Survey. Between 5 and 6% of the population move across counties in the United States each year, according to tax records (Molloy et al., 2011), while 5% of cellphone users in our data move across postcodes in Switzerland in a year (Table A.2 below).

https://doi.org/10.1016/j.jue.2020.103277
Received 8 November 2019; Received in revised form 20 August 2020
Available online 27 August 2020

In practice, gross migration flows are many times larger than net flows, with apparently similar people simultaneously moving in opposite directions (Davis et al., 2016; Monras, 2018). Furthermore, migration flows react slowly, even in the face of large shocks. To help account for these features, moving costs and idiosyncratic location preferences have been added to the classic spatial equilibrium framework (Moretti, 2011; Kline and Moretti, 2014; Diamond, 2016). However, researchers tend to have little information that can give content to the mostly-unobservable individual-location component of residential preferences. Often, this is limited to assigning a distinct status to each person's birthplace or to past locations where they may return (Kennan and Walker, 2011; Diamond, 2016).

In this paper, we document the vital role played by the location of each person's social connections in determining their idiosyncratic preferences for specific locations. Our emphasis on social networks and how they interact with local characteristics is consistent with a spatial equilibrium framework. Precisely if we are close to a spatial equilibrium, common determinants tend to balance out (expensive homes offset high-paying jobs and lake views), and features that are specific to an individual-location pair drive most moves. Instead of treating it as noise, we would like to understand the individual-location component better. Gathering information on a person's network of friends and family helps fill this component with particularly relevant content.

We use information about changes in individuals' neighbourhood of residence and each individual's social network, derived from anonymised cellphone Call Detail Records (CDRs), in combination with demographic and location attributes. The CDRs correspond to all calls made between the universe of customers of a Swiss telecommunications operator (with a 55% national market share in a country with virtually universal cellphone penetration) over the twelve months between June 2015 and May 2016, as well as calls made by these customers to customers of other operators over the same period. These data also include information on each customer's residential address every month between December 2012 and May 2016 as well as on key demographic characteristics. We measure social connections between individuals based on the cellphone calls they make to each other. Individuals can, of course, interact in other ways, such as meeting face-to-face or texting each other. However, as we discuss in detail when describing our data in Section 2, most people use some combination of all three methods to communicate. Calls can be measured more reliably on a large scale than direct encounters. Also, calls are a better indication of close connections and frequent interactions than text messages —particularly between those living far apart. These data enable us to study the role of social networks in residential mobility decisions.

Our analysis shows that taking into account where each person's contacts live doubles our ability to predict who moves and where. Thus, social connections help us understand why similar people make different choices and why the same location attributes have very different effects on them. A significant part of the cost of moving is leaving friends and family behind, and we find that individuals with few local contacts are more likely to change residence. When people move, they strongly prefer places where they already have more contacts living close-by.

We show that the value attached to being close to friends and family accounts for a large fraction of moving costs. Previous studies have found that perceived migration costs are many times larger than the financial costs of moving over a given distance (e.g. Kennan and Walker, 2011; Bayer et al., 2016). Our results suggest that about one-half of the costs that would conventionally be attributed to moving over a certain distance can be accounted for by how that move changes the location relative to the individual's social network. Another way to quantify the importance of contacts is to ask from how much longer individuals would be willing to commute to their current job to be closer to friends and family. We find that someone living right next to their employer would be willing to change residence to a home requiring the average Swiss commute of about half an hour instead, if that meant increasing the share of their contacts who live within 10 minutes by 30 percentage points (equivalent to a 1.3 standard deviation increase in this share).

Three main reasons make contacts matter for residential location choices. First, proximity to contacts is itself valuable and also complements attractive location characteristics. Second, local contacts lower moving costs, for instance, by reducing search frictions when looking for a new home. The third benefit of social connections is that they provide hard-to-find local information that is useful when choosing among alternative locations. In this respect, not only direct connections but also second-order links (friends of friends who are not one's friends) matter, and this finding supports the conclusion that there is an important information channel through which social networks affect residential location choices. Also, contacts who are themselves better connected, as measured by their eigenvector centrality in the overall Swiss network, play a particularly prominent role. We also examine differences across demographic groups. The types of information that matter vary in expected ways with demographics. For instance, individuals aged 25–44 are more likely to move to locations where childcare spots are available if they have contacts there who can tell them about this, while this is irrelevant for those aged 45 and over. Interestingly, as people age, proximity to family gains importance relative to friends.

Studying the role of social networks in residential location choices is complicated by several aspects. We find that people who change residence are more likely to choose a location where they already knew more people close-by. A first obvious concern is that knowing more people in the vicinity of a particular destination may just reflect having lived there before. For this reason, in all our empirical specifications to study the probability of choosing a particular location, we include a return migration indicator, using the information on the individual's prior residential history.

A second concern is that similar individuals are both more likely to be friends and to have similar location preferences. To address the possibility that the importance of local contacts for residential choices reflects such sorting or correlated effects, we begin by controlling for interactions between location characteristics, observable individual characteristics, and individual location history. The empirical importance of social networks is then indicative of whether two individuals with the same demographics and past location history have a different probability of choosing a particular location because one of them knows people close-by and the other one does not. This strategy still leaves open the possibility of sorting on unobservables —perhaps investment bankers are more likely to have other investment bankers as friends as well as to reside in specific postcodes in the country and move predominantly between these. To account for such sorting, we include a full set of origin-destination pair fixed effects at the postcode level, and this only weakens the importance of local contacts by about one-fifth of the baseline magnitude. Finally, we can take advantage of the fact that our data capture actual interactions instead of inferring them from co-location. Thus, we can study the influence of recent movers from the same origin, separating the effect of movers who are among the individual's contacts from the effect of movers who are not. Controlling for strangers who made the same relocation choices leaves the influence of contacts moving from the same origin almost unchanged. These results show that, while relocations across certain location pairs are particularly common, this is not what drives the effects we measure for social networks.

A third concern is that, even if an individual has not lived in a potential destination before, this person may build networks selectively at a place where they are about to move, creating reverse causality.[6] When we study individuals' social networks, we find them to be quite stable up

[6] Whether individuals are more likely to choose locations where they have many pre-existing contacts or locations where they can build new contacts more easily, one might argue that both support our conclusion that social connections provide benefits making settling in a new location easier and more enjoyable. Nevertheless, in our analysis we try to establish a link from pre-existing contacts to the location choice.

to three months before they move. Based on this observation, we build our network measures using exclusively those contacts the individual already had at least four months before moving. We also exclude business phone numbers, to avoid counting calls to a prospective employer or a real estate agent as contacts at a new location.[7] We have also experimented with increasing the gap between the time window in which we characterise the network and the moving date and this has no bearing on our results.

Networks shape many economic decisions and outcomes. See Ioannides (2013) and Jackson et al. (2020) for detailed descriptions of this literature.[8] However, as Topa and Zenou (2015) note in their recent survey of neighbourhood and network effects, “there are very few empirical studies that explicitly test the interactions between the urban space and the social space and their impact on the outcomes of individuals.” This is partly because network ties are often inferred from having two individuals live or work in close proximity or attend the same school (e.g. Bayer et al., 2008; Billings et al., 2019), so that spatial and social proximity cannot be separated. A recent exception is Kim et al. (2020), who study the role of geographical location for social capital.

Some research does collect data on actual network connections, typically using survey techniques. Since such surveys entail high costs, research is often restricted to a few areas or focused on developing countries, where data collection is less expensive (e.g. Alatas et al., 2016). Cellphone CDRs instead provide direct evidence of actual interactions across a vast network. Recent research has also exploited data from online social networks, such as Twitter, Facebook, or LinkedIn.[9] These data are useful to capture alternative channels to transmit information, but online connections are much more weakly related to direct personal interactions than calls (Stopczynski et al., 2014).

Perhaps the two papers closest to ours are Costa et al. (2018) and Koşar et al. (2019). Costa et al. (2018) study the residential location choices of US Civil War veterans and find that after the war they tended to move to a neighbourhood where men from their same war company lived. Veterans appear to have supported one another, as proximity to former comrades raised life expectancy. Koşar et al. (2019) elicit residential mobility and location choice probabilities by presenting nearly 2000 respondents to the New York Fed’s Survey of Consumer Expectations with a series of hypothetical choices. Like we do, they study both the probability of changing residential location and the residential location choice conditional on moving. One of the survey questions asks respondents to imagine a situation where they were forced to move today for at least 3 years to a location 200–500 miles away and had to choose among two locations that differed only in terms of having family and friends move with them or not. They find that proximity to family and friends is the location attribute for which respondents have the highest willingness to pay, amounting to about 30% of annual income (overall non-pecuniary moving costs amount to about 100% of annual income). Other papers study how contacts influence rural-urban migration in a developing country context, without considering the residential location choice among different neighbourhoods or towns (Munshi and Rosenzweig, 2016; Giulietti et al., 2018; Blumenstock et al., 2019).[10]

The remainder of the paper is organised as follows. We begin by describing our data and how we process these in Section 2. Then, in Section 3 we present our estimation approach. Section 4 studies the decision of whether to change residential location or to stay put. Then, conditional on deciding to move, we study the choice among alternative locations in Section 5. The choices of individuals appear to be influenced particularly strongly by contacts who used to be co-residents, and we examine this in Section 6. Distinguishing between those movers across two locations with whom individuals have interacted and those with whom they have not, allows us to show contacts matter over and above shared tastes. In Section 7 we quantify the importance of nearby contacts for residential location choices. Our results highlight that contacts are an important source of information, and focus on this in Section 8. We finally develop a strategy to distinguish friends and family and see how much each group matters and how this varies with age in Section 9. Section 10 concludes.

## 2. Data

### 2.1. Using cellphone calls to capture social interactions

We measure social connections between individuals based on the phone calls they make to each other. Individuals can, of course, interact in other ways, such as meeting face-to-face or exchanging text messages. However, calls are particularly appropriate to study how having social connections who live in different places affects the probability of changing residence and moving closer to them. Users who currently reside in different locations are more likely to talk on the phone than to meet in person or text each other.[11] Interactions between users can also be measured more reliably on a large scale through phone calls than through proxies for direct encounters.[12] Moreover, since most people use some combination of calls, direct encounters, and text messages to commu-

[7] We present results on how social networks affect the probability of changing residence and, conditional on moving, the probability of choosing a particular destination. Individuals who do not feel attached to a place and anticipate moving soon may not bother making many friends locally. Our strategy of using a predefined network of non-business contacts addresses the possibility of building connections in a location to which an individual is about to move, but is much less effective at tackling the possibility of not building contacts locally in anticipation of leaving soon. For this reason, our results on the probability of moving should be interpreted with more caution regarding potential reverse causality than our results on the location choice.

[8] The topics covered include job market referrals and labour outcomes (e.g. Bayer et al., 2008; Beaman and Magruder, 2012; Hellerstein et al., 2014; Brown et al., 2016; Barwick et al., 2019), school performance (e.g. Calvò-Armengol et al., 2009), technology adoption (e.g. Bandeira and Rasul, 2006; Conley and Udry, 2010; Barnejee et al., 2006), crime and incarceration (e.g. Calvò-Armengol and Zenou, 2012; Bhuller et al., 2020; Billings et al., 2019), nest-leaving (Patacchini and Arduini, 2016), and financial market contagion (Kelly and Ó Gráda, 2000) among others.

[9] Bailey et al. (2018) use Facebook data to study a different housing choice, not where to live, but whether to rent or buy. Individuals with Facebook friends in far-away markets with larger house price increases are more likely to transition from renting to owning and to buy large expensive homes.

[10] This last paper also has in common with ours the use of cellphone CDRs, which they use to measure both individuals’ social contacts and to trace whether they are located in urban Kigali or in 27 polygons in rural areas, based on the coverage provided by each of the country’s 30 cellphone towers.

[11] Zignani et al. (2015) explore how calls and text messages relate to physical proximity and find that text message use declines more rapidly with distance, i.e. users who live far away — and can only be together once in a while — are more likely to call than to text each other. It is also worth noting that in Switzerland even the most affordable cellphone plans typically include unlimited calls to all Swiss phone numbers in their flat fee, with plans differentiated primarily based on the amount of data included. Thus, voice calls involve a zero monetary marginal cost.

[12] Direct encounters are usually not observed by researchers but instead inferred from location data. Modern cellphones gather location information from the identifier of the cell tower providing coverage to the user (stored by cellphone operators) and from location data collected by smartphone apps (subsequently purchased, combined, processed and resold by private companies acting as aggregators). Alternatively, Bluetooth technology can be used to track proximity of two cellphone users within a narrower distance, but this usually requires that they install and use a purposely-built app. Stopczynski et al. (2014) issued 1000 Danish university students voluntarily participating in their study with cellphones and an application that used Bluetooth technology to scan for other participants’ devices within an estimated 10-metre range. After merging these data with cellphone records, they find that cellphone calls are a very good predictor of face-to-face contact. The strongest 10% of face-to-face interactions account for 90% of cellphone call ties.

nicate, two people who call each other are very likely to interact more broadly.[13]

### 2.2. Data on telephone communications and individual characteristics

The main dataset used in this paper comprises the anonymised Call Detail Records (CDRs) of all calls originated and/or received by all customers of a large Swiss cellphone operator between June 2015 and May 2016. These include 2.7 million private cellphone lines making 1.8 billion calls over this twelve-month period.

The anonymised CDRs include a hash code that replaces the originating phone number and serves as unique anonymous identifier for this number, a hash code that similarly serves as unique anonymous identifier for the destination phone number, a date and time stamp indicating when the communication was initiated, and the duration of the communication if it was a call. Each hash code identifying a phone number also has associated binary codes indicating whether it is a cellphone or a land line, and whether it belongs to a private or a business customer.

Along with the anonymised CDRs, the operator provided some matched anonymised customer information. This includes the postcode of the billing address, the gender of the customer, a ten-year age bracket (15–24, 25–34, etc.), and the language of correspondence (German, French, Italian, or English). In addition to the monthly postcode of the billing address during the twelve-month calling period, we were provided annual postcode information pre-dating the calling period, starting in January 2012. We use this additional billing address information to differentiate long-term residents, defined as those who have been residing in the same postcode for at least three full years prior to the potential moving date. Note that, since the long-term resident status is based on permanency up until the potential moving date, it can apply to movers as well as stayers.

The anonymity of the operator's customers was guaranteed at all steps of the analysis. We never dealt with or had access to uncensored data. A data security specialist employed by the data provider retrieved the CDRs from the operator's database and anonymised the telephone numbers using a 64-bit hash algorithm. He also removed information on the transmitting cell tower, so that the location of customers at the time of making or receiving calls cannot be traced. The monthly customer information was also censored to include only the aforementioned variables and a hash code to match it with the CDRs. The anonymised data were copied to a fully sealed and encrypted workstation on the operator's premises and we performed all of the analysis on site.

The size of our dataset is large, reflecting the 55% share of the country's cellphone market of our data provider in 2015 (Eidgenössische Kommunikationskommission, 2015). The distribution of cellphone customers in our sample across gender, age, and language groups closely matches that of overall Swiss population as reflected in census data.[14] There are also very strong correlations between our sample and the census in terms of both the number of individuals living in each area and their socio-demographic characteristics at increasingly detailed levels of geographic disaggregation. Even at a very local level, the data is highly representative of the Swiss population, both in terms of its geographic distribution and in terms of its demographic coverage. Tables showing the representativeness of our sample in terms of overall users and of movers are provided in Appendix A.

### 2.3. Measuring residential location and mobility

We assign cellphone customers in our data a residential location based on the postcode of their billing address. This gives us 3152 potential residential locations, each corresponding to a distinct postcode.[15] We measure location characteristics not just at the postcode level (e.g. housing variables), but within given travel times of this postcode (e.g. the share of the individual's contacts reachable in less than 10 min), at the municipal level (2322 units, e.g. for childcare availability), and at the district level (148 units, e.g. crime data).

The billing address is a particularly reliable source of home address information in Switzerland. When private persons residing in Switzerland move, they are legally required to register their new address with the municipality where it is located within 14 days of moving. The Swiss Post office will redirect mail to the new address and also proactively notify at no extra cost the change of address to the companies providing phone service, utilities, etc. on behalf of individuals who have just moved. In addition, Swiss companies can regularly check their customers' addresses against the Swiss Post database to update the billing information of anyone for whom they have an old address, unless the customer has disallowed this. Based on changes in their billing address, we see that 5% of cellphone customers in our data changed their residence to a different postcode between June 2015 and May 2016.

When we compare mobility by cellphone customers in our data with mobility by the Swiss population at large as recorded in the Swiss Post database, we see that the percentage of movers over a twelve-month period is very similar. If we split residential relocations by travel time between the origin and destination postcodes, the distribution of moves is also remarkably close. Both our cellphone data and the Swiss Post data show about 23% of relocations taking place between postcodes separated by up to 10 min of travel time, 32% between postcodes 10–20 min apart, 16% between postcodes 20–30 min apart, 9% between postcodes 30–40 min apart, and the remainder across larger distances. If we correlate residential relocations in both datasets at the postcode level, we also see that our data is remarkably representative of the geographical distribution of moves.

### 2.4. Sample restrictions

We use CDRs mainly to characterise social networks, but not every instance of phone activity reflects a social interaction in a strict sense, so the dataset needs to be filtered beforehand.[16] We centre our analysis on calls between Swiss cellphone numbers belonging to private customers served by our data provider. The reason for focusing on cellphones is that they are almost always used by a single individual and are thus representative of that person's social network. Landlines are instead routinely shared by multiple users and their calls would thus capture overlapping social networks. Excluding cellphone numbers registered to a company is important to ensure that calls reflect a social and not a business interaction. Since we are mainly interested in analysing how the location of social ties affects residential location choices, we need the home address location of caller and callee, so for most of our analysis we rely on intra-operator calls. However, our measures of network topography, such as each individual's eigenvector centrality in the calling network, use both intra-operator and inter-operator calls.

---

[13] Using cellphone records that include information on the transmitting cell tower, Calabrese et al. (2011) find that 93% of cellphone users who call each other have been face-to-face one or more times in the previous year. Remarkably, the figure remains above 90% even for individuals living 100 km apart. Similarly, Wang et al. (2011) show that the frequency of direct encounters between cellphone users is highly correlated with their frequency of calls. Recent survey data show that "phone calls have remained popular in Switzerland despite the onslaught of messaging services" and that most users rely on a combination of calls and messaging, with calls used for more meaningful and complex interactions (Moneyland, 2018).

[14] The only notable difference is that cellphone use is somewhat more prevalent among the very young (ages 24 and under) and somewhat less prevalent among the oldest age group (75 and older), and this is reflected in the age composition of the provider's customer base.

[15] Our set of 3152 postcodes excludes a small number of special codes that are not usable for tracking potential residential locations, such as those assigned to large hospitals.

[16] For a discussion on filtering of cellphone data, see Blondel et al. (2015).

Our starting sample is made up of 2.7 million distinct private cellphone customers making a total of 608 million intra-operator calls to other private cellphone lines. We exclude accidental calls by dropping calls with a duration of less than 10 s. We also drop cellphone numbers that display implausibly low or high monthly usage statistics, with a minimum threshold for the total monthly call duration of one minute and a maximum threshold of 56 h. This removes inactive, or nearly inactive, numbers as well as private lines that may be used for commercial purposes. Finally, we exclude from the analysis customers aged under 15 or over 84 and those for whom information on the residential location and demographic characteristics is unavailable. This yields the final sample size of 2.1 million cellphone customers and 410 million calls.

### 2.5. Defining the social network matrix

Our primary aim is to study how having social ties living in different places affects the probability of changing residence and moving closer to them. Thus, we would like to characterise each individual's social network at the time of a potential move using only calls that reflect a pre-existing social relationship. Excluding calls to and from business numbers already greatly reduces the likelihood that they are made, for instance, to a real estate agent or a school in a prospective new location. However, calls made to private numbers close to the moving date may also reflect an attempt to obtain information or organise details of the move through someone (perhaps a friend of a friend) who is not a pre-existing tie. To more accurately capture first-order social ties, when characterising each individual's social network, we leave a gap between the time window for which the network is computed and the potential moving date being considered.

The choice of time window and gap to the potential moving date used to characterise the network in our baseline specifications is guided by panel (a) of Fig. 1. To produce the figure, for each mover in our sample we calculate how many distinct numbers they call or call them each month from seven months before their relocation date until seven months after their relocation date. We express those as a percentage of the monthly average for that person six to four months prior to their relocation date. The dots represent the mean value across all movers and the bars the standard deviation. We see that individuals start having phone calls with more numbers than usual three months prior to moving, that this number keeps increasing until the moving date, and goes back down and stabilises a couple of months after the move.

Based on panel (a) of Fig. 1, we characterise the social network of an individual on the basis of calls made and received between six and four months before the potential moving date being considered. We observe calls between individuals for the twelve-month period between June 2015 and May 2016. Since we use three months of CDRs to characterise the network and leave a three-month gap to the potential moving date, the first moving date we can consider is December 2015. We study how an individual's decision of whether to relocate then, and if so where, is affected by their social network computed based only on calls made or received by the individual between June and August 2015. This is illustrated in panel (b) of Fig. 1, which shows the six potential moving months that we consider and the three-month window used to compute the social network affecting the decision for each of them, always leaving a three-month gap in between. As a robustness check, we have tried leaving different gaps between the window used to compute the network and the potential moving date and obtained very similar results. Our results are robust to leaving the largest possible gap with our data, eight months.

For each month, we construct an adjacency matrix indicating whether each pair of individuals has called one another. We treat each call as an interaction for both individuals, regardless of who initiated it. Taking into account the residential address of each individual's contacts during that month, we convert the adjacency matrix into a matrix linking individuals to postcodes. Each element $i, j$ of this matrix lists, for that month, how many contacts individual $i$ has spoken with on the phone who resided in postcode $j$ at the time of the call. We then add up these individual-to-postcode matrices over the three-month window six to four months before the potential moving month we are considering.

Since postcodes vary in size, we aggregate contacts at the postcode level into contacts residing within some comparable ring centred on the postcode. Bailey et al. (2020) use Facebook connections between New York City residents to show that geographic distance is an imperfect proxy for social connections, and that actual travel times are more relevant to explain how social ties are formed and maintained. For this reason, we define the rings in terms of travel times. We construct a travel-time matrix providing the time it takes to travel across any two postcodes in Switzerland. These travel times are obtained from https://www.search.ch and correspond to travel by private car under normal traffic conditions. While people may travel by public transport instead, travel times by private car and by public transport are very highly correlated (0.89 correlation for postcode-to-postcode travel times).

Combining the three-month individual-to-postcode matrix with the distance matrix, we calculate our *main measure of the strength of social ties that individual i has in postcode j in month t: the share of all people that individual i spoke with on the phone between months* $t-6$ *and* $t-4$ *who, at the time of the call, resided within 0–10 min travel time of postcode j.* Likewise we calculate the share of each individual's contacts who reside within 10–20 min, 20–30 min, and 30–40 min of each postcode. For robustness, we have also re-estimated our specifications using as the main network variable the number (instead of share) of each individual's contacts who reside within 0–10, 10–20, 20–30, and 30–40 min of each postcode, with very similar results. In all our specifications, we either control for the total number of contacts each individual has (when studying the decision of whether to change residential location), or keep this number constant across options (when studying the choice among alternative locations, conditional on deciding to move). Thus, there is little difference between using the share or the number of contacts within some travel time, although the number variables produce slightly less accurate location choice predictions than the share variables (9.7% instead of 10.2% exact matches at the postcode level). Note that all of these measures are individual specific.

### 2.6. Data on location characteristics

We complement the phone data with variables measuring relevant characteristics for each location. In our estimations, we use location fixed effects to capture the combined impact of everything that makes a location generally attractive or unattractive. Since these location fixed effects absorb the effect of all location characteristics by themselves, the purpose of assembling data on specific location characteristics is to construct individual-location interactions. Thus, when assembling data on location characteristics, we focus on elements that may matter more or less depending on the individual's observable demographic characteristics. We also attach importance to location characteristics that may make having nearby contacts more or less important. The purpose of including these interactions between location characteristics and our network variables is not just to more accurately identify the core effect of contacts on location choices. After all, this can be done with interactions between our network variables and a location fixed effect, an approach that we also implement. Interactions between location characteristics and our network variables allow us to study specific channels through which contacts matter.

For instance, many Swiss neighbourhoods have a very tight housing market and friends and family may greatly help find a new home. We use data on the number of houses and apartments advertised as available to rent or buy on all platforms in the Swiss market in the years 2015 and 2016 for each postcode, obtained from Meta-Sys. We take the average over these years and divide this by the average local housing stock 2015 and 2016, obtained from the Swiss Federal Statistics Office, to compute a relevant measure of housing market tightness.

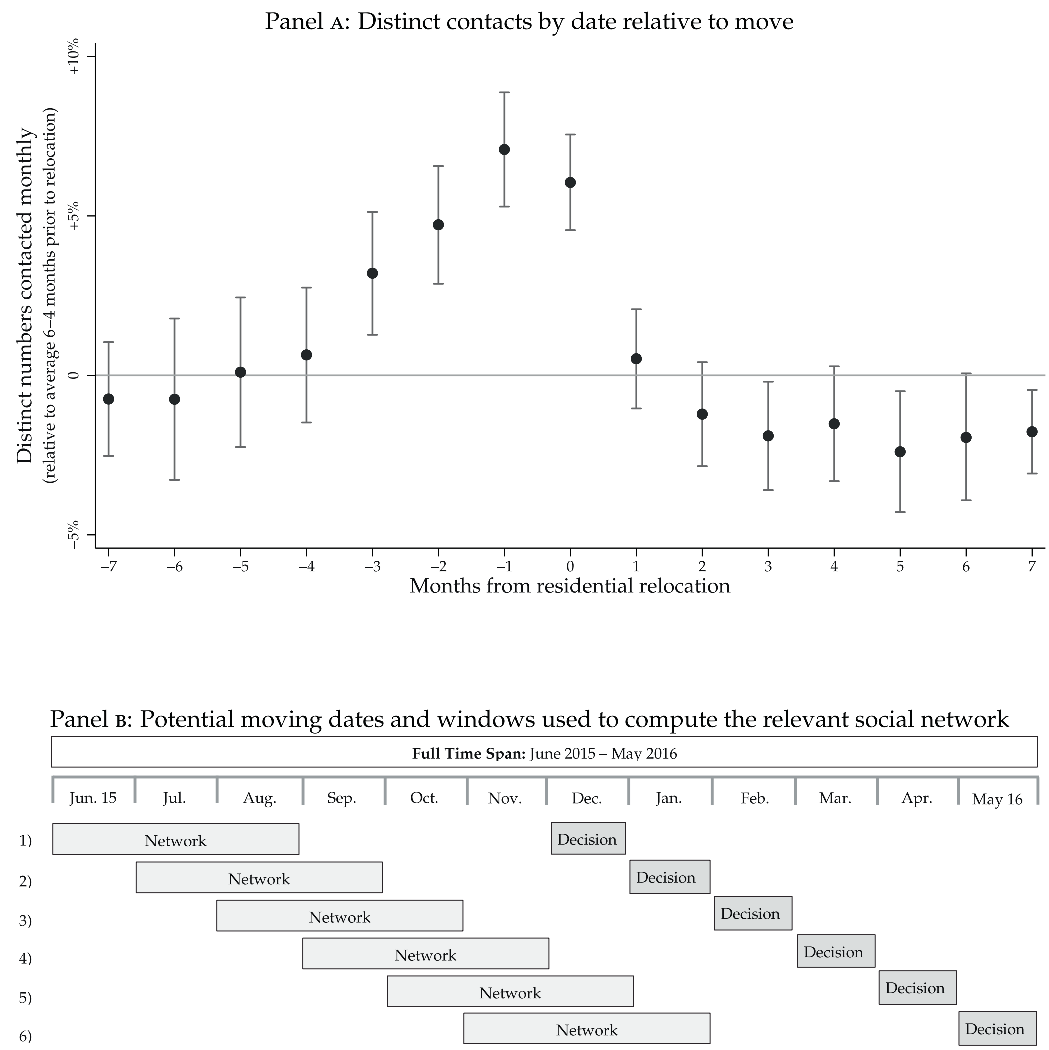


**Fig. 1.** Timing of relocation decisions.

Information on the supply of childcare slots at the local level is not easily available — and this is precisely why we think this will play a very different role depending on whether one has a local contact who can provide information on available slots or not. To get around this, we estimate the number of childcare slots in each municipality based on data about federal subsidies for childcare provided by the Federal Social Insurance Office.[17] The variable we use is the number of childcare slots relative to the local population of children aged 0–14, using census data for the latter.

Crime data is obtained from the Swiss Federal Statistics Office at the district level. This is also the source for data at the postcode level on the share of foreign immigrants, average household size, population density, the local share of home-ownership (as measured by the share of residences inhabited by the owner), and the local income tax burden (defined as income taxes paid to all levels of government by a single earner with an annual income of 100,000 CHF).

We collect detailed information about cultural events using the https://www.guidle.com database. This provides us with the number and type of cultural events by postcode, which we split up into events that target a broad audience and those that cater to a young audience.

Finally, we overlap digitised maps of employment areas, cantons, districts, municipalities, postcodes and majority language areas to assign each postcode to the respective higher level geographical aggregates.

## 3. Framework and estimation

We base our estimation strategy on a two-stage approach. We first analyse the binary decision of *whether* to change residence or not (migration decision). Then, for those who relocate, we study the decision of *where* to move (location decision). Most papers either focus on the decision of whether to migrate (e.g. Finnie, 2004; Blumenstock et al., 2019) or on the location choice conditional on moving (e.g. Schmidheiny, 2006; Agrawal and Foremny, 2019), whereas few papers consider both (Koşar et al., 2019, being an exception). We consider both decisions because networks are likely to matter for attachment to the current place of residence, as well as for factors governing the choice among alternative potential locations. A key advantage of this two-stage approach is that it allows us to separate the costs of moving out from the current location from the costs of moving in to each potential new location.

The indirect utility individual $i$ attains at location $j$ is a function of a component that in turn depends on individual characteristics ($f(\mathbf{X}_i)$), of location characteristics as captured by a location fixed effect ($\lambda_j$), and of

[17] The Federal Social Insurance Office provides subsidies to childcare facilities according to the number of childcare slots. Since virtually all childcare facilities apply for these subsidies, this allows backing out childcare slots at the municipal level. We successfully contrasted the accuracy of these estimates based on information for two cantons where the childcare slots per municipality have been surveyed.

a component that depends on individual-location characteristics ($\pi_{ij}$):

$$V_{ij} = f\left(\mathbf{X}_i\right) + \lambda_j + \pi_{ij} + I_{ij,t-1}\left(\tilde{\lambda}_j + \tilde{\pi}_{ij}\right) + \epsilon_{ij}, \tag{1}$$

where $\epsilon_{ij}$ denotes any unobserved preference components. $I_{ij,t-1}$ is an indicator which is unity if individual $i$ already resided at $j$ in the previous period and zero otherwise, so that $I_{ij,t-1}(\tilde{\lambda}_j + \tilde{\pi}_{ij})$ captures heterogeneity in the location and individual-location components for stayers with $I_{ij,t-1} = 1$ and movers with $I_{ij,t-1} = 0$. For instance, a tight local housing market may be an important drawback for someone moving to a neighbourhood for the first time but much less relevant for someone who has been living there for some time and already has a suitable home. This differentiation between movers and stayers is also consistent with the finding in the migration literature that inflows and outflows respond differently to shocks (Monras, 2018).

We divide the individual-location-specific component into a network component $g(\cdot)$ and another, more standard, component $\mathrm{h}(\cdot)$:

$$\pi_{ij} = g\left(\mathbf{N}_{ij}, \mathbf{N}_{ij}\mathbf{Z}_j, \mathbf{N}_{ij}\mathbf{X}_i\mathbf{Z}_j\right) + h\left(\mathbf{X}_i\mathbf{Z}_j\right). \tag{2}$$

The network component includes a vector $\mathbf{N}_{ij}$ with the shares of individual $i$'s contacts who reside within certain travel times of location $j$. Note that the vector of individual characteristics $\mathbf{X}_i$ in Eq. (1) includes the total number of contacts that individual $i$ has, so the share of contacts variable measures the spatial distribution of contacts controlling for scale. For some specification, we also consider the shares of second-order links within these time intervals and also separate first-order links into subgroups —e.g. recent movers from the same origin or contacts who are particularly central or strong.

In addition to direct network effects, we also study their interaction with location characteristics. The second element of the network component $\mathbf{N}_{ij}\mathbf{Z}_j$ captures the extent to which certain location characteristics, $\mathbf{Z}_j$, are more or less relevant depending on the spatial distribution of the individual's social network. Coming back to the housing example, even for someone moving into a neighbourhood, a tight local housing market will be less of a concern if they have friends and families close-by who can help them find a suitable vacant residence. Since not all relevant characteristics are observable or easily measurable, we also estimate more generic specifications where instead of interactions between network characteristics and observable location characteristics, $\mathbf{N}_{ij}\mathbf{Z}_j$, we include interactions between network characteristics and the location fixed effect $\mathbf{N}_{ij}\lambda_j$. The third element of the network component $\mathbf{N}_{ij}\mathbf{X}_i\mathbf{Z}_j$ allows observable location characteristics to matter differently not only depending on the individual's availability of nearby contacts but also on their demographic characteristics, $\mathbf{X}_i$. For instance, having local contacts who can provide information about whether childcare spots are readily available in a neighbourhood will only be relevant to movers who either have children or are at an age where they may have them soon.

The other individual-location component in Eq. (2), $h(\mathbf{X}_i\mathbf{Z}_j)$, captures interactions between non-network individual characteristics, $\mathbf{X}_i$, and location characteristics, $\mathbf{Z}_j$. A particularly important element here is accounting for whether the individual has previously resided in a given location (an interaction between a location indicator and an indicator for the individual being a former resident). More generally, these non-network individual-location specific elements help us address sorting of heterogeneous individuals to different places. Since this still leaves open the possibility of sorting on unobservables, we complement this with other strategies. In particular, we include a full set of origin-destination pair fixed effects at the postcode level, which accounts for relocations across certain location pairs being particularly common. We also study the influence of recent movers from the same origin, separating the effect of movers who are among the individual's contacts from the effect of movers who are not. Note that, while we use the same vectors, $\mathbf{X}_i$ and $\mathbf{Z}_j$, to denote, respectively, individual and location characteristics in functions $f(.)$, $g(.)$, and $h(.)$, the actual characteristics that matter in each may vary.

### 3.1. Estimating the probability of changing residential location

According to Eq. (1), an individual will decide to migrate if there exists another location that provides higher utility than the current residence $r$:

$$\mathrm{Prob}[\max_{j\neq r}(\lambda_j + \pi_{ij} + \epsilon_{ij}) > \lambda_r + \pi_{ir} + \tilde{\lambda}_r + \tilde{\pi}_{ir} + \epsilon_{ir}]. \tag{3}$$

Put differently, an individual decides to migrate if the utility at the current place of residence $r$ drops below an individual-specific utility threshold given by the best personal alternative. We think of this best personal alternative as follows. Planning a move is a costly process. At any point in time, the individual will not have invested in figuring out every detail about every possible alternative location. Instead, the individual will have information about the distribution of location and individual-location characteristics and individual-location shocks that will allow getting an accurate estimate of the highest utility that would be attainable somewhere else. Provided the difference between that estimate and the utility provided by the current location does not exceed the cost of moving, the individual will stay put.[18] Otherwise, the individual will decide to migrate. We estimate the probability of moving by using linear probability as well as logistic models. As we pool six moving windows (depicted in Fig. 1), we include time fixed effects in all empirical specifications of (3).

### 3.2. Estimating the residential location choice

The second-stage of our approach explores the location choice of those individuals that decided to move. Having decided to move, the individual will gather additional information to figure out the actual values of location and individual-location characteristics and individual-location shocks for specific locations. We estimate the likelihood that a location alternative $k$ provides the highest level of utility among the movers' choice set of locations indexed by $j$, conditional on $j \neq r$:

$$\mathrm{Prob}[\max_{j\neq r}(\lambda_j + \pi_{ij} + \epsilon_{ij})] = \lambda_k + \pi_{ik} + \epsilon_{ik}. \tag{4}$$

As before, the individual-location-specific component includes the shares of the individuals contacts who reside within certain travel times of each potential location. Note that, in the second stage residential location choice, we compare the utility provided by different locations to the same individual. Thus, all individual characteristics are common across all alternatives and no longer appear explicitly in our specifications. One of these individual characteristics is the individual's total number of contacts, so the share of contacts variable still measures the spatial distribution of contacts controlling for scale (where scale is constant across alternatives).[19]

Assuming that $\epsilon_{ij}$ is drawn from an extreme value distribution, the probability that $i$ chooses location $k$ is

$$P_{ik} = \frac{\exp(\lambda_k + \pi_{ik})}{\sum_j \exp(\lambda_j + \pi_{ij})}, \tag{5}$$

[18] This cost of moving is part of what is captured by $\tilde{\lambda}_r + \tilde{\pi}_{ir}$ (the part of the utility individual $i$ gets at location $r$ only because they are already residing there, which they would lose if they moved to another identical location).

[19] The components of Eq. (1) that are interacted with the indicator for the previous location being unchanged, $I_{ij,t-1}$, do not appear in our specifications on the location choice conditional on moving because, by definition, movers are those who change location. Conceptually, this is justified by the assumption that, prior to deciding whether to move, the individual has enough information about the distribution of location and individual-location characteristics and individual-location shocks to make an accurate estimate of the highest utility that would be attainable somewhere else. That is, the individual knows whether they can do better elsewhere, just not exactly where until they invest in gathering more detailed information. This simply rules out that, having decided to move, the individual changes their mind (which, if it happened, we would not be able to observe).

which can be estimated using a conditional logit model.

Analyzing migration and location decisions separately provides a higher degree of expositional clarity compared to an alternative approach estimating (4) based on a sample of movers and non-movers and allowing for $k = r$. In addition, the latter approach would typically involve assuming a common cost of moving for all locations, a frequent simplifying assumption in the literature. Our results suggest this assumption is not supported by the data. In Eq. (1), $\tilde{\lambda}_j$ for $j = r$ captures those moving costs that are common across individuals potentially departing from their current residence in location $r$. If these moving costs were the same across locations, they would be well captured by a constant $\tilde{\lambda}$, so that $\lambda_j + \tilde{\lambda}_j \approx \lambda_j + \tilde{\lambda}$. The location fixed effects of the first stage of our approach ($\lambda_r + \tilde{\lambda}_r$) would then be highly correlated with the location fixed effects of the second stage ($\lambda_j$ for $j = r$). Instead, we find a low correlation between the location fixed-effects of both stages, which indicates that moving costs are heterogeneous across locations. Finally, the two-stage approach allows for heterogeneous effects of individual-location interactions, network variables and network-location interactions in the migration and location decisions (i.e. $\pi_{ij} + \tilde{\pi}_{ij} \neq \pi_{ij}$). This seems quite relevant as, for instance, in the housing example above: one may expect that the scarcity of house vacancies interacted with the strength of the network at a moving destination matters for the location choice of movers (since friends may help to find a new home) whereas it seem unlikely to be relevant for the migration choice (since, at that point, the person is already settled in a home).

In principle, all 3152 postcodes are available as location alternatives. However, with 47,214 movers this yields about 150 million observations which is computationally not feasible in a non-linear model. We address this issue using two alternative approaches. First we estimate a linear probability model where the dependent variable is an indicator $d_{ik}$, which takes value one if individual $i$ has chosen location alternative $k$ and zero otherwise. We regress this on location fixed effects $\lambda_k$, defined at the postcode level, as well as individual-location variables $\pi_{ik}$:

$$d_{ik} = \lambda_k + \pi_{ik} + \epsilon_{ij}. \tag{6}$$

Second, we explore reasonable restrictions of the choice set based on network information. It turns out that over 95% of all movers go to a location where six to four months before relocating they already had at least one contact residing within 40 min. Based on this, we use all location alternatives with a contact within a radius of 40 min. Focusing on such locations reduces the choice set for the average individual to about 540 postcodes. We also consider 100 additional random locations, and adjust the estimates applying sample weights (Manski and Lerman, 1977; Cosslett, 1981). Where computationally feasible, we also consider all 3152 postcodes and, when we do this, we get almost identical estimates.

## 4. The decision to relocate

We begin our empirical analysis by studying each individual's decision about whether to move away from their current residential location. We conjecture that a significant part of the cost of moving is leaving friends and family behind. This argument underlies the provisions for family-based international migration of many countries. It has also been used to explain why closely-connected communities are more resilient in the face of economic shocks or natural disasters (UNE, 2018). In an internal migration context, Coate and Mangum (2019) argue that most of the recent decline in mobility within the United States is due to increasingly tight social ties making people more rooted in what used to be high-mobility locations (where rootedness is proxied in their analysis by parents and children sharing birthplace). They also suggest that asymmetries in the cost of gathering information at the current versus alternative locations may be important. While our focus is on current mobility decisions as opposed to time trends, cellphone CDRs give us a direct measure of how rooted individuals are to their current location through the intensity of their local ties.

In Table 1, we estimate the probability of changing residence to a different postcode as a function of individual, location, and individual-location characteristics (where location is the current place of residence), compared to some individual-specific outside option. Our focus is on the social network structure for each individual, and we characterise this through a combination of individual-location measures and individual measures.

The main individual-location social network measure is the share of the person's contacts residing within a certain travel time from their current home. As explained in Section 2 and illustrated in panel B of Fig. 1, for each potential moving date, the individual's contacts are all those with whom the individual established at least one intra-operator phone call (undirected, in the sense that it could have been initiated by either party) in the period six to four months before that date. In addition, we also consider the share of second-order links (i.e. friends of friends who are not one's friends) located within a certain travel time from their current residence. The main individual social network measure is the total number of contacts the individual has, which is conventionally named degree centrality. This degree centrality variable controls for the scale of the individual's network, while the share variables measure its spatial distribution.

We also include individual-location and individual characteristics other than our network measures. As additional individual-location characteristics, we include an indicator for whether the individual is a long-term resident in their current location, in the sense of having resided in the same postcode for at least three full years, and also an indicator for whether the individual shares the local majority language. As additional individual characteristics, we include language, age, and gender. The estimation pools data for six possible moving dates, one for each month between December 2015 and May 2016, so we also include month fixed effects.

Columns (1) to (3) estimate the probability of changing residential location using a linear probability model, while columns (4) to (6) do so using a logistic model. In Column (1), our key variable of interest is the share of the individual's contacts located within 10 min travel time from their current residence. As expected, the estimated coefficient is negative and statistically significant. The estimated coefficient indicates that the magnitude of the effect is large: an increase of one standard deviation (i.e. 0.289) in the share of the individual's contacts that are located within 10 minutes reduces the probability of moving from the average 4.8% to 3.9% (calculated as $4.8 - 0.289 \times (-3.008)$).[20]

An individual with few local contacts may be someone whose social network is mostly located elsewhere, but also someone who is not very sociable. To account for sociability separately, we include in our specifications degree centrality, defined as the individual's total number of contacts. The coefficient on degree centrality is positive and statistically significant. An increase of one standard deviation (9.944 additional contacts) raises the probability of moving from the average 4.8% to 4.9%. We interpret this as evidence that more sociable and connected individuals, leaving aside the spatial distribution of their contacts, are slightly more mobile.

Shifting attention to non-network individual variables, we see that individuals who share the local majority language and long-term residents (those who have been residing at the current location for at least three years prior to the potential moving date) are less likely to move.

In Column (2), we add further network variables. Looking at the coefficients for the share of each individual's contacts who reside within 10–20 min, 20–30 min, and 30–40 min of the current postcode, we see that they are all negative and statistically significant. In terms of magni-

[20] See Appendix B for descriptive statistics of the independent variables. For the decision about whether to change residential location, the relevant statistics are those measured for all individuals at their original place of residence, with the mean and standard deviation shown in, respectively, columns (1) and (2) of Table B.1.

**Table 1**
Probability of changing residential location.

| Dep. var.: Probability of changing residential location | Linear probability model | | | Logit | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Share of contacts | | | | | | |
| 0–10 min | −3.008*** | −3.596*** | −3.199*** | −1.426*** | −1.606*** | −1.518*** |
| | (0.039) | (0.050) | (0.061) | (0.019) | (0.022) | (0.029) |
| 10–20 min | | −1.450*** | −1.327*** | | −0.478*** | −0.458*** |
| | | (0.059) | (0.060) | | (0.023) | (0.024) |
| 20–30 min | | −0.717*** | −0.671*** | | −0.211*** | −0.204*** |
| | | (0.072) | (0.072) | | (0.028) | (0.028) |
| 30–40 min | | −0.365*** | −0.343*** | | −0.151** | −0.148*** |
| | | (0.086) | (0.086) | | (0.034) | (0.034) |
| Share of 2nd-order contacts 0–10 min | | | −1.025*** | | | −0.217*** |
| | | | (0.094) | | | (0.049) |
| Total number of contacts | 0.010*** | 0.011*** | 0.011*** | 0.005*** | 0.005*** | 0.005*** |
| | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) |
| Long-term resident | −0.815*** | −0.798*** | −0.795*** | −0.327*** | −0.321*** | −0.321*** |
| | (0.022) | (0.022) | (0.022) | (0.009) | (0.009) | (0.009) |
| Speaks same language as majority | −0.508*** | −0.388*** | −0.379*** | −0.123*** | −0.083** | −0.081* |
| | (0.076) | (0.077) | (0.077) | (0.027) | (0.027) | (0.027) |
| Language, age, gender | Yes | Yes | Yes | Yes | Yes | Yes |
| Location fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.016 | 0.016 | 0.017 | – | – | – |
| Pseudo $R^2$ | – | – | – | 0.180 | 0.180 | 0.181 |
| $N$ | 2,136,093 | 2,136,093 | 2,136,093 | 2,136,093 | 2,136,093 | 2,136,093 |

*Notes*: Dependent variable is expressed as a percentage in the linear probability model. Location fixed effects defined at the postcode level in columns (1)–(3) and at the employment region level in columns (4)–(6). The pseudo $R^2$ in columns (4)–(6) is calculated following McKelvey and Zavoina (1975) and reflects the proportion of the variance of the dependent variable that is explained by the covariates. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

tude, the deterrent effect on mobility of having a larger share of contacts within 30–40 min of the current residence (instead of the baseline over 40 min away) is about one-half as large as the deterrent effect of having them within 20–30 min of the current residence. In turn the effect of having a larger share of contacts within 20–30 min is about one-half as large of the effect of having them within 10–20 min, which in turn is about one-half as large of the effect of having them within 0–10 min.

Our results in Table 1 indicate that an individual who has a smaller share of social contacts living nearby is less rooted locally and more likely to move away. However, we cannot disentangle whether the individual has tried to make friends locally and not been able to do so, or whether this person feels less attached to the location for other reasons and, anticipating an upcoming move, has not made much of an effort to establish local ties. While we include as controls an indicator for long-term residency up until the potential move date and the total number on contacts, it is still possible that having a small share of contacts who are local is partly endogenous to the desire to move.

To address somewhat such endogeneity concerns and also to isolate the role of contacts in providing useful information, in column (3) we incorporate the share of the individual's second-order contacts located within 10 min travel time from their current residence. These second-order contacts are friends of the individual's friends that have not interacted with the individual directly. Since these second-order links are friendships established by someone else, they are exogenous to the individual. However, they can still provide useful information indirectly. For instance, an individual living in Bern's Lnggasse neighbourhood may have few local friends, but some of this person's friends in another location, say Biel, may in turn have friends who also live in Lnggasse. Although the individual has never talked to these neighbours directly, he or she may still get advice about a new local restaurant or child-care facility or a job referral from these unknown neighbours indirectly through their common friend in Biel. We see in column (3) that the share of second-order links located within 0–10 min from the individual's current location also makes a change of residence less likely. This suggests that networks matter greatly for information gathering and that information is an important determinant of residential location choices.[21]

The results for the logistic model of columns (4) to (6) match those of the linear probability model. Note that, while the coefficients are not directly comparable, calculating the effect of an increase of one standard deviation in the share of the individual's contacts that are located within 10 min gives a reduction in the probability of moving from the average 4.6% to 2.9%, a larger effect than estimated in the linear probability model. We have also re-estimated the same specifications measuring the proximity of the individual to contacts with the number, instead of the share, of contacts within a given travel time. Results (not reported in Table 1) remain almost the same.

## 5. The residential location choice

We now turn the second step of our analysis, where we study —conditional on moving— the role of a person's social network in choosing their new residential location. Our data includes 47,214 individuals who move to a different postcode over the six possible moving months considered, December 2015 to May 2016. For each of these movers, we examine how their choice of a new location is influenced by the net-

[21] First-order links combine an information advantage and also the direct enjoyment of interacting with them. In the case of second-order links, there is not a direct benefit of interactions because what distinguishes second-order from first-order links is that for the former direct interactions have not taken place. At the same time, note that we cannot interpret the difference between the coefficients on first-order and second-order links within 10 min as measuring the effect on mobility of the direct enjoyment of interaction with close-by contacts; this difference also likely reflects the greater effectiveness of gathering information directly (from first-order links) versus indirectly (second-order links). Note also that, while we control explicitly for long-term residency, the importance of second-order links may partly reflect that people who are more established are both less likely to move and to have made friends who themselves have more local connections.

**Table 2**
Residential location choice.

| Dep. var.: Probability of choosing a location conditional on moving | | | | | | |
|---|---|---|---|---|---|---|
| | Linear probability model | | | | | Cond. logit |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Share of contacts | | | | | | |
| 0–10 min | 9.259*** | 8.855*** | 9.085*** | 9.084*** | 9.044*** | 6.263*** |
| | (0.015) | (0.015) | (0.015) | (0.014) | (0.015) | (0.027) |
| 10–20 min | 0.949*** | 0.909*** | | | | |
| | (0.007) | (0.007) | | | | |
| 20–30 min | 0.049*** | 0.045*** | | | | |
| | (0.005) | (0.005) | | | | |
| 30–40 min | −0.038*** | −0.038*** | | | | |
| | (0.004) | (0.004) | | | | |
| Return migration | | 15.318*** | 15.386*** | 15.383*** | | 2.765*** |
| | | (0.036) | (0.036) | (0.036) | | (0.026) |
| Individual × location controls | Yes | Yes | Yes | No | Yes | Yes |
| Individual controls × location f. e. | No | No | No | Yes | No | No |
| Location-specific return migration | No | No | No | No | Yes | No |
| Location fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.024 | 0.030 | 0.030 | 0.030 | 0.030 | – |
| Pseudo $R^2$ | – | – | – | – | – | 0.207 |
| $N$ | 25,555,189 | 25,555,189 | 25,555,189 | 25,555,189 | 25,555,189 | 25,342,595 |

*Notes*: Dependent variable is expressed as a percentage in the linear probability model. Location fixed effects defined at the postcode level in columns (1)–(5) and at the employment region level in column (6). Return migration indicates the individual was a resident at the same location at an earlier time in 2012–2015. Individual × location controls are an indicator for the potential new location being in the same employment region as the current residence, an indicator for the individual's preferred language being the local majority language, an interaction between an indicator for the individual's preferred language being English and the local share of foreign immigrants, an interaction between an indicator for the individual having multiple cellphone numbers on the same bill and the local average household size, an interaction between six age-group indicators and local population density, an interaction between six age-group indicators and the local share of homeowners, and an interaction between six age-group indicators and the local tax burden. Individual controls × location fixed effects uses as individual controls age group and gender indicators. The pseudo $R^2$ in column (6) is calculated following McFadden (1973). ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

work of contacts with whom the individual spoke on their cellphone six to four months before the moving date.

The dependent variable in Table 2 is the probability that an individual mover chooses a specific location over all other alternatives. As before, residential locations are defined as each of the 3152 postcodes in Switzerland. Estimation with such a large set is computationally challenging. However, our data show that over 95% of all movers go to a location where six to four months before relocating they already had at least one contact residing within 40 min. Focusing on such locations reduces the choice set for the average individual to about 540 postcodes. Where feasible, we have estimated our specifications in three different ways: considering only locations with at least one pre-existing contact within 40 min, considering in addition 100 random locations, and considering all 3152 postcodes. All yield essentially identical results.[22]

The table is now estimated using only movers and identifies coefficients based solely on variation for a given individual. For this reason, the specifications no longer include individual characteristics, only location and individual-location characteristics. Throughout our regressions analysing residential location choices, instead of considering specific location characteristics (e.g. housing prices, local tax rates, geography, climate, etc.) one by one, we absorb all of them into a location fixed effect. This is because our focus is on understanding the importance for residential location choices of each individual's social network, and in particular of how this network is distributed across space. Regarding individual-location characteristics, the main individual-location social network measures are again the share of the person's contacts residing within a certain travel time from their current home. We also include individual-location characteristics other than our network measures, and we describe these as we incorporate them into to our empirical specifications.

Columns (1) to (5) estimate the probability of choosing a specific new residential location using a linear probability model, while column (6) does so using a conditional logit. We begin with the linear probability model and will focus on this for much of the analysis. The main reason is that in the linear probability model it is computationally feasible to include a location fixed effects for each of the 3152 postcodes in Switzerland, whereas in the conditional logit we can only include location fixed effects for each of the country's 16 employment regions.[23]

The results in column (1) indicate that having pre-existing contacts within a short travel distance of a given postcode increases the likelihood of choosing that particular postcode when changing residence. The estimated coefficient for the share of contacts within 10 min travel time is positive and significant.[24] It indicates that having 10% more local contacts within this range increases the base probability of choosing that location relative to all others by almost one percentage point ($0.10 \times 9.257$) —a substantial effect given that there are more than 3000 postcodes to choose from and that the average individual has pre-existing contacts within 40 min in around 540 postcodes.

Turning to the coefficients on the share of contacts located further away, we see that it is mostly very local contacts that matter (those

[22] For Table 2, considering all locations merely increases computing time. For the more demanding estimations further below, estimation considering all 3152 postcodes as relevant alternatives for every individual becomes infeasible. Thus, in the remainder of the main text we restrict the choice set for each individual to postcodes where they have at least one contact within 40 min.

[23] The second, more standard, advantage of the linear probability model is its interpretability, since the estimated coefficients can be read as the change in the probability of moving (where this probability is expressed as a percentage in our tables) for a one-unit change of the independent variable of interest, holding everything else constant. The main disadvantage, of course, is that the linear probability model does not constrain probabilities to the unit interval.

[24] Standard errors are likely understated due to spatial correlations within and across nearby postal codes. While the large choice set each individual faces makes dealing with this issue computationally infeasible, the standard errors are so small that we believe a correction for spatial correlation would not change the significance of our estimates in any meaningful way.

located within 0–10 min driving distance from the possible new residence). Comparing the coefficients for the share of the individual's contacts located within the different travel windows, we see that contacts located within 10–20 min of a potential new residential location only matter one-tenth as much as those located within 0–10 min and those located further matter even less. Note this decline with distance is more pronounced for the choice of where to move than for the choice of whether to move — in Table 1, we saw that in that context contacts located a further 10 min away mattered about one-half as much.

We also control for other individual-location characteristics, mostly meant to capture how well the individual and the location match.[25] The results (coefficients not reported in the table) show that choosing a given new postcode is more likely if this is located within the same employment area as the postcode of previous residence, presumably because this allows to change home without changing jobs. Postcodes where the majority language matches the individual's preferred language are also more likely to be chosen. Also postcodes where the location characteristics match well with observable individual characteristics have a higher probability of attracting movers. Including these individual-location controls increases the explanatory power of our model relative to estimating the same specification without them, but does not alter the importance of the variables characterising where contacts are located. This large set of individual-location characteristics controls to some extent for the sorting of individuals with certain observable characteristics into the same type of neighbourhoods. This matters because similar individuals are also more likely to be friends). In the following section, we develop further strategies to deal with sorting.

As discussed in the introduction, the available data limits much of the literature to assigning a special role only to the individual's birth location or to locations where the individual has lived before when trying to identify the attachment of an individual to a particular location. One may worry that our measures regarding the presence of a large share of the individual's contacts close to a potential new location may just reflect that the individual is returning to an earlier residential location. In column (2), we show that this is not the case by adding an indicator for return migration. This return migration indicator takes value one if the individual was a resident at the same location at an earlier time, as captured by the billing address history prior to our calling data period. The corresponding coefficient is positive and statistically significant, but bringing this indicator into the regression has almost no effect on our variables characterising the spatial distribution of the individual's social network: the first four coefficients in column (2) are very similar to those in column (1). Thus, while return migration is frequent, accounting for this almost does not affect the importance of where contacts are located for the choice of where to move.

Given the very fast decay with distance in the importance of contacts for residential location choices, in column (3), we re-estimate the specification of column (2) considering only the share of contacts within 10 min. We will also do this for subsequent specifications.

Individual observable characteristics may matter not only in relation to the locations characteristics we have included in our specification, but also in relation to others that we have not considered or cannot measure. With this in mind, in column (4) we re-estimate the specification of column (3), but interacting individual characteristics with postcode fixed effects. This makes no difference to our key coefficients.

Since different locations may draw previous residents back to them to a different extent, in column (5), instead of having a common return migration indicator, we include a separate return migration indicator for each of the 3152 postcodes. This improves the fit but leaves the key coefficients essentially unchanged.[26]

Finally, column (6) replicates the estimation of column (3) using a conditional logit instead of a linear probability model, finding comparable results. We have also re-estimated Table 2 using the number of contacts (instead of the share of contacts) within a given travel time. Results (not reported) remain very similar, both under the linear probability model and under the conditional logit. We prefer the share of local contacts as main measure for an individual's proximity to social contacts, since it yields a slightly better model fit than the number of contacts within a given travel time.

## 6. Chain mobility and sorting

When people change residence, they often follow in the footsteps of other recent movers from the same origin. This is a particularly well-known phenomenon in an international migration context, where foreign immigrants tend to locate, at least initially, in ethnic enclaves. In fact, this behaviour has served as a basis for numerous studies on the consequences of immigration on labor markets, which exploit variation across local markets in immigrant flows (see Dustmann et al., 2016, for a review). Following Altonji and Card (1991) and Card (2001), it is common to use a migration-networks instrument to account for the endogenous sorting of migrants across locations. This strategy instruments actual migrant flows at the local level with flows by immigrant group at the national level weighted by the initial stock of each group at the local level. The relevance of this instrument is based precisely on the fact that past stocks of immigrants in specific locations are good predictors of future flows.

Persistent flows of individuals from the same origin to the same destination may reflect chain migration, as defined by MacDonald and MacDonald (1964, p. 82): "a movement in which prospective migrants learn of opportunities, are provided with transportation, and have initial accommodation and employment arranged by means of primary social relationships with previous migrants." However, such flows may also reflect sorting or correlated effects: individuals with similar characteristics tend to prefer living in similar locations, so when they move it is more likely that they were living in the same location before and also that they end up living in the same location again —even if they have never met.

Since network links are typically inferred from past co-location, it is usually difficult to separate actual network effects from sorting. One strategy is to use interactions of individual characteristics with location characteristics as controls to account for sorting on observables, as we have done in Section 5. However, this mitigates but does not eliminate the possibility that sorting on unobservables is important.

We have seen that potential new locations with more pre-existing contacts nearby are more likely to be chosen. We are worried that this key result may partly reflect a tendency of similar individuals (who are more likely to be friends) to relocate across the same postcodes more

[25] We include an indicator for whether the potential new residential postcode is within the same employment area as the current residence, an indicator for whether the postcode is within a language region that corresponds to the pre-existing billing language of the customer, an interaction between an indicator for the individual's preferred language of correspondence being English and the share of foreign immigrants in the potential new location (since foreigners may be more likely to choose locations where many other foreigners also live), an interaction between an indicator for the individual having multiple cellphone numbers on the same bill and the local average household size (since families may be more likely to choose the same neighbourhoods where there are many other families), an interaction between six age-group indicators and local population density (since individuals may be more or less likely to locate in central versus suburban locations at different stages of life), an interaction between six age-group indicators and the local share of homeowners (younger individuals are less likely to be homeowners and may be more likely to choose neighbourhoods where rentals predominate), and an interaction between six age-group indicators and the local tax burden (this is the closest to sorting by income we can capture with our data). These individual-location controls are in addition to the return migration indicator discussed separately.

[26] Individuals may also be more or less likely to return to a previous location depending on their age and gender. When, in addition to having a return migration indicator, we interact this indicator with the individual's gender and age bracket, results (not reported) remain essentially the same.

**Table 3**
Chain mobility.

| Dep. var.: Probability of choosing a location conditional on moving | | | | | |
|---|---|---|---|---|---|
| | Linear probability model | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| Share of contacts 0–10 min | 6.931*** | 8.518*** | 8.166*** | 8.702*** | |
| | (0.181) | (0.015) | (0.017) | (0.015) | |
| & recent movers | | | −0.459*** | | |
| | | | (0.051) | | |
| & from same origin | | | 9.107*** | | |
| | | | (0.094) | | |
| Share non-contact movers from same origin | | 88.881*** | 88.160*** | | |
| | | (0.470) | (0.470) | | |
| Return migration | 13.525*** | 15.275*** | 15.261*** | 16.640*** | 15.336*** |
| | (0.039) | (0.036) | (0.036) | (0.036) | (0.036) |
| Moving distance | | | | −0.208*** | −0.438*** |
| | | | | (0.002) | (0.002) |
| Individual × location controls | Yes | Yes | Yes | Yes | Yes |
| Location fixed effects | No | Yes | Yes | Yes | Yes |
| Origin-destination pair fixed effects | Yes | No | No | No | No |
| $R^2$ | 0.147 | 0.031 | 0.031 | 0.030 | 0.017 |
| $N$ | 25,555,189 | 25,555,189 | 25,555,189 | 25,555,189 | 25,555,189 |

*Notes*: Dependent variable is expressed as a percentage. Location fixed effects defined at the postcode level. Recent movers captures the additional effect of contacts who, in addition to residing in any postcodes that can be reached by car within 10 min, moved there between January 2013 and three months prior. From same origin further restricts these to those who moved from within 10 min driving distance of where the individual is also moving. Share non-contact movers from same origin considers those individuals who moved between January 2013 and three months prior from within 10 min driving distance of where the individual is also moving and who are not one of their contacts and then calculates what share of these chose a postcode that can be reached by car within 10 min. Individual × location controls as in Table 2. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

generally, irrespective of network effects. A possible strategy to deal with this concern is to include a full set of origin-destination pair fixed effects, which capture that moves across certain postcode pairs may be particularly likely. We do this in column (1) of Table 3. Compared with our specification in column (3) of Table 2, the coefficient on share of the individual's contacts within 10 min travel time of a potential destination postcode falls by about one-fifth but is still large and highly statistically significant. Note this coefficient is now identified only on the basis of variation in the distribution of contacts within origin-destination pair (essentially comparing two individuals with identical demographics and observed location history who are departing from the same postcode and differ only in terms of the share of their contacts within 10 min of a common potential destination).

Yet another strategy to address concerns about possible correlated effects is to take advantage of the richness of our data, which allows us to separate network effects from co-location. Column (2) in Table 3 replicates our specification in column (3) of Table 2 controlling for the share of non-contact movers from the same origin to a given location. To construct this variable, we consider those individuals who moved between January 2013 and three months prior to the individual's moving date from within 10 min driving distance of where the individual is also departing. We then identify who among these are not one of the individual's contacts, and finally calculate what share of these non-contact movers from the same origin chose a postcode that can be reached within 10 min of each potential new location. We see that this variable is positive and significant, indicating that certain origin-destination pairs are particularly likely to be shared even by individuals who are not socially connected to one another. However, remarkably, the coefficient of share of contacts 0–10 min remains almost identical. This is evidence that our key results regarding the importance of networks do not merely reflect the sorting of similar people into similar places.

In column (3) of Table 3 we further disentangle the effect of local contacts on relocation decisions. Specifically, we allow these local contacts to matter differently depending on whether they arrived at the potential new location recently or they have instead been there for some time. For those who arrived recently, we also differentiate between those who moved from the same location from which the individual is now departing and those who moved from a different location. To this effect, we use three related variables. First, as before, "share of contacts 0–10 min" is the share of all of the individual's contacts who can be reached by car within 10 min of the potential new location. The second variable is "share of contacts 0–10 min & recent movers," calculated as the share of all of the individual's contacts who, in addition to residing in any postcodes that can be reached by car within 10 min, moved there between January 2013 and three months prior. The third variable "share of contacts 0–10 min & recent movers & from same origin" further restricts this share to those who moved from within 10 min driving distance of where the individual is also departing.

With all three variables simultaneously in the regression, the effect of those among the individual's contacts who are long-term residents of a potential new location on the probability that this location is chosen corresponds to the coefficient on "share of contacts 0–10 min." The effect of contacts who have only moved there recently but from a different origin than the individual corresponds to the sum of the coefficients on "share of contacts 0–10 min" and "share of contacts 0–10 min & recent movers." Finally, the effect of contacts who used to live close to the individual and moved to the new potential location recently corresponds to the sum of all three coefficients. Looking at the signs and magnitudes, we see that local contacts who used to live close to the individual and moved recently to a new location increase the probability of choosing that new location by twice as much contact who are long-term residents of the new location. We conjecture that the individual may get more useful information from these social ties who recently completed the same move, either because they have been in more direct contact recently or because they just went through the same process and have more relevant tips to share. Contacts who arrived recently from an entirely different destination do not matter very differently than long-term residents at the new location.[27]

[27] Once gain, conditional logit specifications (not reported), when feasible to estimate, give a similar message as the linear probability model reported in Table 3.

**Table 4**
Predictive power.

| | Location FE, indiv. × location controls & share local contacts | Location FE & indiv. × location controls |
|---|---|---|
| | (1) | (2) |
| Correct predictions at postcode level | 10.2% | 5.1% |
| Correct predictions at district level | 43.8% | 18.2% |
| $R^2$ linear probability model | 0.030 | 0.015 |
| Pseudo $R^2$ conditional logit | 0.253 | 0.116 |

*Notes*: Correct predictions calculated as the share of movers for whom the location with the highest estimated probability of being chosen matches their actual choice, following Domencich and McFadden (1975). Column (1) corresponds to the estimation of Table 2 column (3). Column (2) corresponds to the same estimation as column (1) without the share of contacts 0–10 min variable.

We comment the results in columns (4) and (5) of Table 3 in the following section.

## 7. Quantifying the importance of nearby contacts for residential location choices

Our results show that taking into account where an individual's contacts live helps us understand that person's choice of residential relocation, but how much does this matter in practice? In this section, we offer three approaches to quantify the importance of nearby contacts for residential location choices. First, we check how much our ability to predict residential relocation choices improves with information about each individual's social network. Second, we estimate what fraction of the cost of relocating over a given distance can be accounted for by changes in proximity to family and friends resulting from the move. Third, we compare choice alternatives that require different degrees of commuting and offer different proximity to local contacts to estimate how much longer would people be willing to commute to work in order to be closer to family and friends.

To get a better idea of importance of nearby contacts for residential location choices, in Table 4 we evaluate the predictive power of our estimations, with and without network characteristics. Following Domencich and McFadden (1975), we compute the percentage of correct predictions from each specification as the percentage of movers for whom the postcode with the highest estimated probability of being chosen matches their actual chosen postcode. Column (1) corresponds to our specification of Table 2 column (3), estimating the probability that a particular individual chooses a specific postcode among the 3152 possibilities available on the basis of postcode fixed effects and individual-location characteristics including the share of contacts within 0–10 min travel time. This specification can guess the exact postcode to where 10.2% of movers relocate. Column (2) in Table 4 corresponds to the exact same specification, removing only the share of contacts within 0–10 min travel time, and this makes the percentage of correct predictions at the postcode level drop by one-half to 5.1%.

Since guessing the exact postcode chosen may be excessively demanding, in the second row of Table 4 we check the accuracy of our predicted choices at the district level (148 units). We do so similarly, by computing the percentage of movers for whom the postcode with the highest estimated probability of being chosen is located in the same district as their actual chosen postcode. When we include the share of local contacts, our estimation correctly predicts relocations at the district level for 43.8% of movers. Once again, if we exclude the share of local contacts, this percentage drops by more than one-half, to 18.2%.

Note both the specifications in column (1) and column (2) include postcode fixed-effects (absorbing all characteristics of each location that may make it more or less attractive to the population at large), as well as a full set of interactions between individual characteristics and location characteristics (capturing the extent to which a location with certain characteristics may be particularly attractive to individuals with certain demographics). Thus, the specification in column (2) corresponds to a relatively standard and complete residential location choice model. Compared with such a standard model, taking into account how many contacts each individual has in close proximity to each location in column (1) doubles our ability to predict where individuals relocate.

Prior research on residential mobility has estimated very large migration costs that increase rapidly with distance (see, e.g., Greenwood, 1997; Kennan and Walker, 2011). A second way to quantify the importance of nearby contacts for residential location choices is to estimate what share of the cost of migrating over larger distances is driven by ending up further away from family and friends. If we draw concentric circles (in travel-distance space) around an individual's current residential location, there will be multiple potential new residential postcodes at any given travel distance, but they will differ in terms of how many contacts this person has within 10 min of each potential destination. Thus, our data allows us to separate the effect of moving distance from the effect of spatial separation from friends and family. In column (4) of Table 3, we add the distance between the origin and the potential destination postcodes (measured by natural logarithm of the travel time by road under normal traffic conditions) to our specification of Table 2 column (3). The coefficient on this distance variable is negative and highly statistically significant (point estimate of −0.208 with standard error 0.002). However, the share of contacts within 0–10 min travel time only diminishes slightly from 9.085 to 8.702. If we then take out the share of contacts within 0–10 min travel time but leave the distance between the origin and the potential destination postcodes in the specification, the coefficient on the latter variable more than doubles to −0.438. This implies that more than half of the deterring effect of distance on choosing a new residence is driven by the greater separation from pre-existing contacts that a more distant move would typically entail.

A third way to quantify the importance of local contacts is to calculate what additional costs an individual is willing to incur in order to live in a location that is closer to family members and friends. In our estimations of the probability of choosing a specific location, location fixed-effects absorb whatever costs are common across individuals, including housing prices. In contrast, the commuting costs associated with a given residential location will differ across individuals depending on their work location. While the information on each individual's job location is not available to us for this article, we are able to use an estimate from Büchel and Ehrlich (2021) for each individual in the same sample that we use of how long it would take that individual to commute to their current job from each postcode. To compare the cost of separation from family and friends with the cost of commuting, we begin from our estimation of Table 3 column (1). Recall that this includes a full set of origin-destination pair fixed effects. We now add to this specification an indicator variable for whether a potential new residential postcode is not within walking distance of the individual's current work location and an interaction of this indicator with the natural logarithm of the time it would take to commute by road from that residential postcode to the current work location. We are essentially comparing two individuals with identical demographics and observed location history who are departing from the same postcode and differ in terms of the share of their contacts within 10 min of a common potential destination and in terms of the length of commute to their current job this new location would entail. This allows us to compare the cost of being far from friends and family with the cost of commuting.[28] Consider an individual who

[28] This comparison is a specific case of a more general trade-off between moving closer to career opportunities and losing social ties. See, for instance, Wahba and Zenou (2012), who model the trade-off individuals face when deciding whether to migrate abroad between accumulating human capital, which facilitates becoming an entrepreneur once they return, and losing social capital.

currently experiences the average commute of 26 min.[29] According to our results, moving to an otherwise identical location within walking distance of the individual's current job —thus avoiding commuting— generates the same utility gain as increasing the share of local contacts by 30 percentage points. This is equivalent to a 1.3 standard deviation increase in the share of contacts within 10 min (see Table B.2, where the average is 20% and the standard deviation 22.6%).

Our results in this section show that the spatial distribution of individuals' social contacts is a quantitatively important factor determining their residential location choices. Using information about each individual's social network doubles our ability to predict their choice of residential relocation. About one-half of the costs that would conventionally be attributed to moving over a certain distance can be accounted for by how that move changes the location of an individual's home relative to their social network. And living in a location where the share of contacts is 30 percentage points higher is valued as highly as being able to avoid the average Swiss commute of about 30 min by residing right next to one's workplace.

## 8. The role of information

A key reason why already knowing people in a prospective neighbourhood matters so much when deciding where to move is that local contacts can provide useful information. Some characteristics of a location (e.g. local tax rates) are public information that is easy to obtain simply through a web search. Other characteristics (e.g. whether a location is a good place to raise kids or whether a location has had a recent uptick in crime) are more difficult to observe from far away. As a result, there is an informational asymmetry between areas where the individual who is considering moving there knows people who are likely to have and transmit this information and areas where the individual knows no-one.

In column (1) of Table 5, we explicitly consider this possibility by incorporating into our benchmark specification in column (2) of Table 3 an interaction between the share of the individual's contacts who live within 10 min travel time of each potential destination location and a location fixed effect. These interaction terms are estimated using the iterative procedure of De la Roca and Puga (2017). As in previous specifications, we define location fixed effects at the postcode level and centre them at zero. Thus, a positive value for a given postcode indicates this location has a set of features that make it more broadly attractive than average. In contrast, a negative value indicates below-average attractiveness. The positive and statistically significant coefficient on the interaction term confirms that having pre-existing social contacts in a location makes moving to that location more likely if the location is particularly attractive. It instead makes moving to that location less likely if the location is particularly unattractive.

Since social contacts are an essential source of information regarding a potential new residential location, their importance is likely to vary depending on how much information these contacts have and how close the individual's relationship to them is. The information that an individual can obtain from their social network depends on the quality of their contacts (more or less central) and also on the intensity of the ties (stronger or weaker) (see Ioannides, 2013; Giulietti et al., 2018). In columns (2) to (4) of Table 5 we re-estimate the specification of column (1), now exploring the position of contacts in the network and the intensity of the links.

Relative to column (1), in column (2) we add "share of contacts 0–10 min & central." This new variable corresponds to the share of the individual's contacts who, in addition to residing within 10 min of the potential new postcode, are in the top 10% in terms of eigenvector centrality in the overall Swiss network (Bonacich, 1972). Eigenvector centrality assigns relative scores to all nodes in the network based on the idea that a node is more important when it is better connected to other important nodes. Since the specification still includes "share of contacts 0–10 min," the coefficient on "share of contacts 0–10 min & central" captures the additional effect of local contacts who are particularly central, showing they are crucial drivers of location choices. We also interact this new network variable with location fixed effects. The positive coefficient on this interaction indicates that central contacts are particularly influential in driving movers towards attractive locations and away from unattractive locations, as captured by postcode fixed effects.

Instead of centrality, columns (3) and (4) consider two measures of link strength between the mover and each of their contacts: the combined duration of calls in column (3) and the frequency duration of calls in column (4). Again, we include both the new network variable and its interaction with location fixed effects. We see that weaker links are a particularly important determinant of migration decisions and a relevant source of information. This finding suggests that when an individual is considering a new location, they get in touch with acquaintances who live close-by even if they are people who they usually do not talk with long or often.[30]

In addition to getting in touch with weaker contacts, movers can also gather information indirectly from friends of friends. One may ask a friend about available houses in her neighbourhood and they may not know of a suitable one but can ask their friends and come back with suggestions. In column (5) of Table 5, we add to our benchmark specification in column (2) of Table 3 the share of the individual's second-order contacts located within 10 min travel time of the potential new location. These second-order contacts are friends of the individual's friends that have not interacted with the individual directly. As in previous columns, we also interact this new network variable with location fixed effects.

The coefficients on first-order and second-order links in column (5) are similar in magnitude. Note, however, this does not imply that a friend of a friend is as useful a source of information as a direct friend. The number of second-order links is far higher than the number of first-order links (by a factor of 16, see Table B.1 in the appendix). Thus, on average, it takes 16 times as many local links to increase the share of second-order contacts within 10 min compared with the share of first-order contacts. In addition, the positive interaction term between first-order contacts and location fixed-effects may reflect the information that first-order contacts provide about the general attractiveness of a location, but also that attractive locations are even more enjoyable in the company of local friends and family. In contrast, by construction, second-order contacts have not interacted with the individual directly, so arguably capture a pure information channel.

To further isolate the role of information provided by social contacts in residential location choices, as well as to isolate specific types of information that matter, in column (6) of Table 5 we interact our primary network variable with relevant measures of local characteristics. The strategy of interacting network variables with location fixed effects used in previous columns captures any characteristics of a location that make it particularly attractive or unattractive to the population at large. However, identifying specific types of information that matter is also of interest. Moreover, since some information will matter differently to various demographic groups, exploring this heterogeneity can give us additional confidence that we are capturing an information channel.

There are at least three relevant types of information that one can gather through contacts. First, friends and family who are already living near a potential new home can provide information before mov-

[29] The average commute of 26 min in our data is very similar to the one obtained in the official commuting survey of 31 min (see https://www.bfs.admin.ch/bfs/en/home/statistics/mobility-transport/passenger-transport/commuting.html).

[30] Our findings that movers tend to follow people they know who migrated recently from the same origin (Table 3) but that weak links are also relevant (Table 5) is in line with the results of Giulietti et al. (2018) regarding rural-urban migration decisions in China. They suggest that contacts who migrated recently provide more direct support and help in settling in while weak ties are relevant for information gathering.

**Table 5**
The role of information.

| Dep. var.: Probability of choosing a location conditional on moving | | | | | | |
|---|---|---|---|---|---|---|
| | Linear probability model | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Share of contacts 0–10 min | 6.620*** (0.016) | 6.394*** (0.017) | 7.198*** (0.019) | 6.887*** (0.019) | 5.083*** (0.009) | 9.038*** (0.019) |
| Share of contacts 0–10 min | | | | | | |
| × location fixed effect | 13.601*** (0.032) | 12.239*** (0.034) | 14.577*** (0.039) | 14.265*** (0.039) | 6.949*** (0.091) | |
| × childcare slots | | | | | | 6.724*** (0.243) |
| × recent crimes | | | | | | −1.603*** (0.014) |
| × cultural events | | | | | | 0.042*** (0.001) |
| × housing turnover | | | | | | −238.758*** (1.623) |
| Share of contacts 0–10 min | | | | | | |
| & central | | 2.992*** (0.052) | | | | |
| & central × loc. fixed effect | | 6.655*** (0.115) | | | | |
| & strong (duration) | | | −1.610*** (0.038) | | | |
| & strong (duration) × loc. f. e. | | | −5.967*** (0.081) | | | |
| & strong (frequency) | | | | −0.570*** (0.039) | | |
| & strong (frequency) × loc. f. e. | | | | −4.836*** (0.081) | | |
| Share of 2nd-order contacts 0–10 min | | | | | 4.420*** (0.037) | |
| Share of 2nd-order contacts 0–10 min × location fixed effect | | | | | 8.991*** (0.066) | |
| Share non-contact movers from same origin | 80.333*** (0.475) | 88.199*** (0.469) | 88.142*** (0.469) | 89.057*** (0.469) | 74.755*** (0.469) | 78.734*** (0.478) |
| Return migration | 14.603*** (0.036) | 14.668*** (0.036) | 14.626*** (0.036) | 14.656*** (0.036) | 14.540*** (0.036) | 15.091*** (0.036) |
| Individual × location controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Location fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.038 | 0.038 | 0.038 | 0.037 | 0.039 | 0.033 |
| $N$ | 25,555,189 | 25,555,189 | 25,555,189 | 25,555,189 | 25,555,189 | 25,538,167 |

*Notes*: Dependent variable is expressed as a percentage. Location fixed effects defined at the postcode level. Central contacts are those in the top 10% in terms of eigenvector centrality in the overall Swiss network. Strong (duration/frequency) contacts are those in the top 10% in terms of total call duration/frequency in the individual's contact network. Location fixed effects defined at the postcode level. All local characteristics are centred at zero. Share non-contact movers from same origin considers those individuals who moved between January 2013 and three months prior from within 10 min driving distance of where the individual is also moving and who are not one of their contacts and then calculates what share of these chose a postcode that can be reached by car within 10 min. Individual × location controls as in Table 2. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

ing that will help rank a prospective neighbourhood above others. Another relevant type of information concerns advice that one may wish to gather through the social network regularly after moving. For instance, a neighbourhood may feature a variety of cultural events or have a trendy nightlife scene. However, to fully take advantage of these amenities, it is useful to know other locals who can share tips of where to go, or who may even join in. Finally, information gathered through friends may help alleviate frictions in search markets. In particular, many Swiss neighbourhoods have very tight housing markets. Given that houses and apartments for rent or purchase are often taken as soon as they go on the market, it becomes extremely useful to garner information about suitable available units through local contacts who may have heard about them through the grapevine, perhaps even before they are advertised.[31]

Column (6) of Table 5 shows all three types of information matter. Starting with characteristics difficult to observe from far away, a first example is childcare availability. Information on the supply of childcare slots at the local level is not easily available. Recall from Section 2 that we got around this by estimating the number of childcare slots in each municipality based on data about federal subsidies for childcare. Note that here the key issue for prospective residents is not to get a spot if they are available. The process for assigning available slots is open and straightforward, so having local contacts will not help get ahead of the queue. The key issue is knowing how easy it is to get a childcare spot. The variable we use is the number of childcare slots relative to the local child population. As expected, we find that the interaction term between the share of contacts living within 10 min and the childcare slots to pupil ratio is positive and significant.[32] The local availability of childcare will be relevant only to people with children or at an age where they may have children soon. When we estimate the same specification separately for individuals aged 25–44 and those aged 45 and over (not

[31] For example, 31% of people found their current residence through their social network, according to the Migration Survey of the Swiss Canton of Basel-Stadt.

[32] Note that the childcare slots to pupil ratio and other local characteristics do not appear uninteracted in the regression because we include postcode fixed effects that will absorb these.

**Table 6**
Relative importance of friends and family by age.

| | All | | Age 25–34 | | Age 35–54 | | Age ≥ 55 | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Share family contacts within 0–10 min | 2.694*** | 2.394*** | 2.378*** | 2.176*** | 3.171*** | 2.485*** | 3.791*** | 2.854*** |
| | (0.019) | (0.019) | (0.027) | (0.028) | (0.045) | (0.045) | (0.075) | (0.074) |
| Share friend contacts within 0–10 min | 7.803*** | 5.695*** | 8.101*** | 5.735*** | 7.565*** | 5.222*** | 6.727*** | 4.734*** |
| | (0.034) | (0.035) | (0.051) | (0.0522) | (0.078) | (0.079) | (0.129) | (0.126) |
| Share family contacts 0–10 min × location fixed effects | | 5.560*** | | 4.190*** | | 6.456*** | | 6.077*** |
| | | (0.046) | | (0.107) | | (0.086) | | (0.087) |
| Share friend contacts 0–10 min × location fixed effects | | 15.540*** | | 16.054*** | | 17.893*** | | 12.420*** |
| | | (0.077 | | (0.580) | | (0.144) | | (0.148) |
| Share non-contact movers from same origin | 70.514*** | 68.555*** | 74.100*** | 73.369*** | 66.314*** | 58.932*** | 65.190*** | 48.915*** |
| | (0.856) | (0.849) | (1.322) | (1.311) | (1.959) | (1.930) | (3.131) | (3.028) |
| Return migration | 15.684*** | 14.411*** | 13.993*** | 12.877*** | 13.628*** | 12.645*** | 13.608*** | 12.522*** |
| | (0.069) | (0.068) | (0.093) | (0.092) | (0.194) | (0.191) | (0.350) | (0.339) |
| Individual × location controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Location fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.041 | 0.056 | 0.040 | 0.055 | 0.037 | 0.067 | 0.041 | 0.104 |
| $N$ | 6,235,721 | 6,235,721 | 2,793,125 | 2,793,125 | 1,067,625 | 1,067,625 | 383,510 | 383,510 |

*Notes*: All columns estimated using a linear probability model, with the dependent variable expressed as a percentage. Location fixed effects defined at the postcode level. Individual × location controls as in Table 2. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

reported in Table 5), we find that having local contacts who can provide information about childcare availability only matters for people whose age makes them more likely to have children now or soon.

We also consider the local prevalence of crime. While violent crimes are rare in Switzerland, other felonies and misdemeanours, such as home burglaries are more prevalent. These are often committed by itinerant crime groups, and as a result high and low crime rate areas change relatively quickly. While getting past crime statistics is relatively straightforward, obtaining information about more recent spurts of crime is complicated unless people you know tell you about current episodes. The interaction term between the share of contacts living within 10 min and the recent local prevalence of crime is negative and significant. This indicates that individuals are less likely to move to a high-crime location if they know someone locally who has warned them about the recent trend.

Turning to information that may be useful after moving, we now consider a measurable example of trendy amenities. The interaction term between the share of contacts living within 10 min and a measure of local cultural events in the period we study is positive and significant. Contacts seem to matter in terms of being able to exchange information about the quality and location of amenities in the neighbourhood and possibly also in terms of enjoying them together. We have also experimented separating local cultural events into those likely to appeal to a young audience and those likely to appeal to a broad target audience. Our results (not reported in Table 5) indicate that broadly-targeted events have a positive effect in combination with local contacts for the two age groups considered, 15–24 and 35 and over. However, events targeted at a younger audience have a positive effect on younger people and a negative effect on older ones (perhaps younger people learn from their contacts about how cool a DJ session is, while older people learn from their contacts about how unpleasantly noisy this was).

The final interaction looks explicitly at the extent to which local contacts can alleviate frictions in the housing market. For each postcode, we add up all the houses and apartments advertised as available to rent or buy on all platforms in the Swiss market in the years 2015 and 2016. We take the average over these years and divide this by the average local housing stock 2015 and 2016 to compute a relevant measure of housing turnover. The interaction term between the share of contacts living within 10 min and this measure of housing turnover is negative and significant. This indicates that postcodes with lower house turnover, where it is more difficult to find a home, are more likely to be chosen if one has local contacts who can alleviate the search frictions.[33]

## 9. Friends and family

It is plausible that the importance of the person's contacts in deciding where to live could be different depending on whether the contact is a friend or a family member. The anonymisation process undergone by the phone records obviously means that we cannot observe which contacts are family and which are friends. However, from the structure of calls between two nodes who call each other (each node being a hash code corresponding to an anonymised cellphone number) and the rest of the network, in combination with the age brackets and gender for each node, we can try to infer whether these nodes are more likely to be connected by a family relationship or by friendship. The process is described in detail in Appendix C. Given the inherent measurement errors in detecting family ties using calling patterns, the results that follow should clearly be regarded merely as suggestive.

Table 6 gives the results for the estimations of the importance of friends and family. Column (1) includes the results when splitting the share of contacts within 10 min into friends and family. Both are positive and statistically significant. Column (2) adds interactions between these contact shares variables and a postcode fixed effect, which are also positive and significant. Thus, as would be expected, both friends and family matter for residential location decisions. In terms of magnitude, note that the average person has many more friends than family members so the larger coefficient on the share of friends variable should not be interpreted as implying that a person influences location decisions more if they are a friends rather than a family member. Instead, it

[33] In addition to housing market tightness, we have also explored labour market tightness. Unfortunately, data on job vacancies is only available at a very aggregate geographical level. The local unemployment rate is available at the municipality level, but is a rough measure of labour market tightness. Nevertheless, when we include an interaction between the local unemployment rate and our primary network variable, we find a positive and significant coefficient. This suggests that having local contacts makes it more likely to choose a postcode within a municipality with higher unemployment. That said, unemployment in Switzerland is very low and much less spatially heterogeneous than housing market tightness. See Barwick et al. (2019) for an analysis of job referrals using cellphone data.

says that on average people give more weight to where their friends are concentrated than to where their family is concentrated. However, this changes as people age. In columns (3) to (8) we repeat the estimation of columns (1) and (2) separately. In columns (3) to (8) we estimate the same equations as columns (1) and (2) but we separate the movers depending on their ages (25–34, 35–54 and older than 54). The results indicate that as people age, proximity to family gains importance relative to proximity to friends. This effect is particularly pronounced for the older group of movers.

## 10. Conclusions

In this paper, we examine the role of a person's social network on the decision of changing her residence and on choosing a new location where to live. We use data on actual interactions, as measured by phone calls between individual cellphone users, in combination with accurate and frequent data on residential location and demographic and location attributes. We organise our estimation strategy in two steps. First, we analyse the effect of the social network on the probability that an individual moves to a new residential location. Results indicate that people whose contacts are more concentrated close to their current residence are less likely to move. We further find that the friends of the person's friends also help to keep them attached to their current location and that more sociable individuals are slightly more mobile. However, distance matters, and for every additional 10 min of travel time required to reach contacts, their importance is slashed by one-half.

In the second step, and conditional on deciding to move, we study the role of the person's social network on her new residential choice among alternative locations. The evidence indicates that the prior presence of local contacts increases the probability of choosing a location. For the specific choice of location, distance matters even more, with those located within 10 min having an effect at least an order of magnitude greater than the rest.

Knowing people in a prospective new neighbourhood matters so much partly because they can provide useful information. We show that, in the context of choosing a residential location, three types of information gathering are important. Having sufficient information prior to the move helps rank a prospective neighbourhood above others. This is particularly important for information that is hard to obtain other than from people with local knowledge (e.g. the local availability of childcare or recent crime spurts). A second aspect to information gathering concerns advice that one may wish to gather through the social network regularly after moving. For instance, a neighbourhood may feature a variety of cultural events or have a trendy nightlife scene. However, to fully take advantage of this it is useful to know other locals who can share tips of where to go or even join in. A third aspect to information gathering concerns searching in markets subject to frictions. In the Swiss context, these friction are most relevant when looking for a new home. We show that knowing locals in neighbourhoods where gathering these three types of information are particularly relevant strongly influences location choices. Not only direct contacts, but also friends of friends who are not ones's friends matter greatly for information gathering.

Our findings show that very different types of contacts affect residential location choices for complementary reasons. Direct friends, in addition to providing information and reducing frictions, also have an important role due to the enjoyment of direct interactions with them. Friends of friends have a particularly strong role for information gathering and weaker direct links also matter on this respect. Movers tend to follow people they know who migrated recently from the same origin and can help them settle at the new location. This pattern of chain mobility does not merely reflect a tendency of similar individuals (more likely to be friends) to relocate across the same postcodes more generally. We show that movers from the same origin to a given location who the individual knows personally continue to matter just as much if we control for movers from the same origin to a given location who are not part of the individual's personal network. When distinguishing between the influence of friends and family, we find that both matter but proximity to family gains importance with age.

The secular decline in the propensity to move across locations within the United States and many countries in Europe and elsewhere is often seen as reflecting substantial frictions that should be reduced. Our results suggest that the reluctance to move and the idiosyncratic pickiness in choosing a new residence may reflect the relevance of social networks. Connections make people more rooted in specific locations and also create important asymmetries in the cost of gathering information. To the extent that people derive utility from being close to family and friends, it is sensible that they trade off this proximity against the advantages of alternative locations. At the same time, insofar informational asymmetries are important, making information more readily available could create significant welfare gains and provide more equitable access to localised opportunities and amenities.

## A. Sample representativeness

**Table A1**
Sample representativeness.

| | Sample | Census | Correlation sample-census at the level of | | |
|---|---|---|---|---|---|
| | | | Employment areas | Districts | Municipalities |
| | (1) | (2) | (3) | (4) | (5) |
| Individuals | $2.1 \times 10^6$ | $6.7 \times 10^6$ | 0.99 | 0.98 | 0.99 |
| Female | 48.47% | 50.43% | 0.98 | 0.98 | 0.99 |
| Average age | 43.70 | 46.61 | — | — | — |
| *Age groups* | | | | | |
| 15–24 | 19.39% | 13.96% | 0.97 | 0.96 | 0.98 |
| 25–34 | 19.06% | 16.57% | 0.97 | 0.96 | 0.98 |
| 35–44 | 15.42% | 17.31% | 0.98 | 0.98 | 0.99 |
| 45–54 | 19.30% | 19.15% | 0.99 | 0.99 | 0.99 |
| 55–64 | 15.14% | 14.63% | 0.99 | 0.99 | 0.99 |
| 65–74 | 9.54% | 11.41% | 0.98 | 0.98 | 0.99 |
| 75–84 | 2.15% | 6.98% | 0.92 | 0.93 | 0.97 |
| *Main Language* | | | | | |
| German | 68.90% | 63.45% | 0.99 | 0.98 | 0.99 |
| French | 26.33% | 20.61% | 0.99 | 0.99 | 0.99 |
| Italian | 4.12% | 6.37% | 0.95 | 0.97 | 0.95 |
| English | 0.65% | — | — | — | — |
| Other | — | 9.49% | — | — | — |

*Notes*: All data on both cellphone users and census population are for individuals aged 15–84.

**Table A2**
Movers representativeness.

| | Sample (1) | Postal data (2) |
|---|---|---|
| Movers across postcodes as % of population | 4.95 | 4.17 |
| % of movers by distance | | |
| 0–10 min | 22.81 | 23.29 |
| 10–20 min | 32.47 | 31.61 |
| 20–30 min | 16.47 | 15.72 |
| 30–40 min | 8.94 | 8.91 |
| >40 min | 19.31 | 20.45 |
| *Correlation with sample* | | |
| Movers by origin postcode | | 0.96 |
| Movers by destination postcode | | 0.97 |
| Movers by origin-destination postocodes | | 0.78 |

*Notes*: Column (1) reports moves based on changes in the postcode of the billing address for cellphone users in our sample for June 2015-May 2016. Column (2) reports moves based on data on address changes recorded by Swiss Post for January-December 2014.

## B. Descriptive statistics

**Table B1**
Descriptive statistics at original residence.

| | All individuals | | Movers | |
|---|---|---|---|---|
| | Mean (1) | Std. Dev. (2) | Mean (3) | Std. Dev. (4) |
| Share of contacts | | | | |
| 0–10 min | 0.385 | 0.289 | 0.281 | 0.248 |
| 10–20 min | 0.211 | 0.220 | 0.221 | 0.213 |
| 20–30 min | 0.124 | 0.170 | 0.145 | 0.172 |
| 30–40 min | 0.079 | 0.137 | 0.096 | 0.143 |
| Degree centrality (total number of contacts) | 10.057 | 9.944 | 11.281 | 9.718 |
| Share of 2nd-order contacts 0–10 min | 0.179 | 0.161 | 0.140 | 0.132 |
| Total number of 2nd-order contacts | 157.736 | 221.132 | 177.280 | 210.611 |
| Long-term resident | 0.651 | – | 0.499 | – |
| Speaks same language as majority | 0.961 | – | 0.955 | – |
| Total number of calls | 75.837 | 100.737 | 97.949 | 111.651 |
| Total call duration (min) | 274.137 | 451.054 | 392.317 | 542.743 |

*Notes*: All variables computed over a three-month window between six and four months before each potential moving month from December 2015 to May 2016 and averaged for each individual over all six potential moving months.

**Table B2**
Descriptive statistics for movers.

| | At new location | | Mean across locations | |
|---|---|---|---|---|
| | Mean (1) | Std. Dev. (2) | Mean (3) | Std. Dev. (4) |
| Share of contacts | | | | |
| 0–10 min | 0.200 | 0.226 | 0.002 | 0.025 |
| 10–20 min | 0.228 | 0.228 | 0.009 | 0.059 |
| 20–30 min | 0.153 | 0.185 | 0.019 | 0.086 |
| 30–40 min | 0.106 | 0.158 | 0.030 | 0.109 |
| Share of 2nd-order contacts 0–10 min | 0.120 | 0.126 | 0.002 | 0.016 |
| Share family contacts 0–10 min | 0.299 | 0.420 | 0.002 | 0.043 |
| Share friend contacts 0–10 min | 0.211 | 0.221 | 0.002 | 0.025 |
| Same employment area | 0.798 | – | 0.082 | – |
| Return migration | 0.051 | – | 0.000 | – |
| Speaks same language as majority | 0.945 | – | 0.522 | – |
| Share non-contact movers from same origin | 0.004 | 0.009 | 0.000 | 0.001 |

*Notes*: All variables computed over a three-month window between six and four months before the moving month.

**Table B3**
Local characteristics.

| | Mean | Std.Dev. | Min | Max |
|---|---|---|---|---|
| **Interaction variables** | | | | |
| Ln (Crime) | 7.529 | 1.212 | 3.694 | 10.617 |
| Housing turnover | 0.015 | 0.011 | 0 | 0.100 |
| Child care slots per child | 0.032 | 0.060 | 0 | 0.696 |
| Number of events marked as highlight | 10.628 | 77.957 | 0 | 2747 |
| Number of events with young target group | 22.447 | 158.908 | 0 | 4014 |
| Number of events without age specific target group | 28.613 | 123.666 | 0 | 3185 |
| **Controls** | | | | |
| Main Language | | | | |
| *German* | 0.645 | – | 0 | 1 |
| *French* | 0.258 | – | 0 | 1 |
| *Italian* | 0.089 | – | 0 | 1 |
| Population density | 488.393 | 1037.905 | 0.112 | 11975 |
| Share of migrants | 16.752 | 9.862 | 0.019 | 60.630 |
| Share of homeowners | 46.088 | 21.788 | 0 | 95.522 |
| Income tax burden | 14.200 | 2.152 | 5.618 | 18.747 |
| Avg. age | 36.493 | 2.357 | 29.430 | 65.000 |
| Avg. household size | 2.347 | 0.243 | 1.180 | 3.236 |

## C. Inferring family ties

In Section 9, we analyse how the relative importance of friends and family for location choices varies over the life cycle. We exploit the structure of calls and socio-demographic information to infer whether contacts are close relatives or friends. In particular, we employ the following algorithm:

1. We extract the call matrices for four three-months periods, i.e. June 2015–August 2015, September 2015–November 2015, December 2015–February 2016, and March 2016–May 2016.
2. Links between customer pairs occurring in less than 3 out of the 4 quarters are dropped.
3. Based on the remaining links and sociodemographic information from the billing data, we assign customers to families. As illustrated in Fig. C.1, we identify six different types of potential family clusters, which we order along the following hierarchy:

**(I)** *Full quad:* We look for pairs of parent-nodes, which we require to be of opposite sex and whose age has to lie within a range of 15 years. The two parent-nodes also need to interact with at least two children that are 20 to 40 years younger. If we observe a complete set of links between the two parent-nodes and the two (or more) children nodes we label the group as full quad family.

**(II)** *Quad with missing parent-parent link:* Among all customers not belonging to a full quad family, we look for parent-nodes that interact with at least two children. If we observe a complete set of

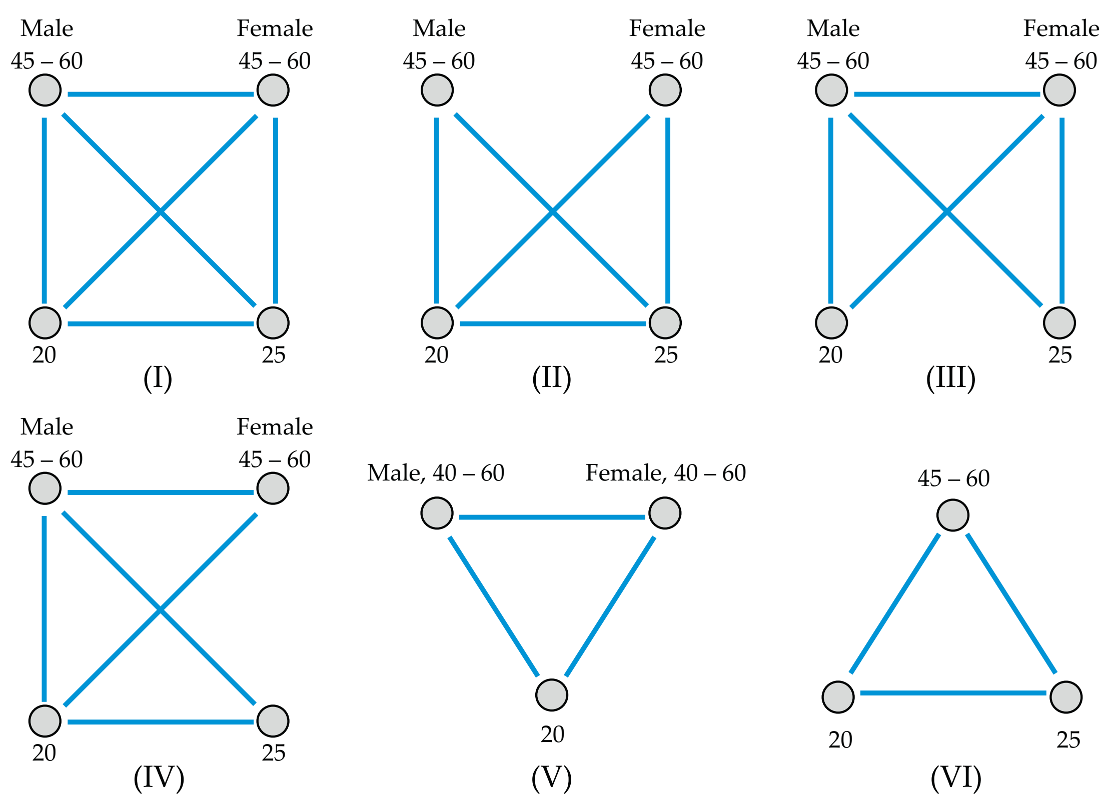


**Fig. C.1.** Hierarchy of Family call patterns, quads & triangles.

links between all four (or more) nodes except between the parents we label the group as quad family with a missing parent-parent link.

**(III)** *Quad with missing child-child link:* Among all customers not belonging to a quad family of type (I) or (II), we look for parent-nodes, that interact with at least two children. If we observe a complete set of links between all four (or more) nodes except between the children we label the group as quad family with a missing child-child link.

**(IV)** *Quad with missing parent-child link:* Among all customers not belonging to a quad family of type (I), (II) or (III), we look for parent-nodes that interact with children. If we observe a complete set of links between all four (or more) nodes except between the one parent and one child we label the group as quad family with a missing child-parent link.

**(V)** *Two parents + one child:* Among all customers not belonging to a quad family of type (I), (II), (III) or (IV) we look for parent-nodes that interact with one child. If we observe a complete set of links between all three nodes we label the group as triangle family with two parents and one child.

**(VI)** *One parent + two children:* Among all customers not belonging to a family of type (I), (II), (III), (IV) or (V) we look for two children that interact with one parent. If we observe a complete set of links between all three nodes we label the group as triangle family with two children and one parent.

4. Once we have assigned customers to the six different types of family clusters as described in step 3, we merge them into families with up to three generations: grandparents, parents, and children.
5. All phone interactions between mobile phone customers that do not belong to the same family clusters are labelled as interactions between friends.

## CRediT authorship contribution statement

**Konstantin Büchel:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Maximilian V. Ehrlich:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Diego Puga:** Conceptualization, Methodology, Formal analysis, Writing - original draft. **Elisabet Viladecans-Marsal:** Conceptualization, Methodology, Formal analysis, Writing - original draft.

## References


Agrawal, D.R., Foremny, D., 2019. Relocation of the rich: migration in response to tax rate changes from Spanish reforms. Rev. Econ. Stat. 101 (2), 214–232.

Alatas, V., Barnejee, A., Chandrasekha, A.G., Hanna, R., Olken, B.A., 2016. Network structure and the aggregation of information: theory and evidence from indonesia. Am. Econ. Rev. 106 (7), 1663–1704.

Altonji, J.G., Card, D., 1991. The effects of immigration on the labor market outcomes of less-skilled natives. In: Abowd, J.M., Freeman, R.B. (Eds.), Immigration, Trade and the Labor Market. Chicago University Press, Chicago, IL, pp. 201–234.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., 2018. The economic effects of social networks: evidence from the housing market. J. Polit. Econ. 126 (6), 2224–2276.

Bailey, M., Farrell, P., Kuchler, T., Stroebel, J., 2020. Social connectedness in urban areas. J. Urban Econ. 118, 103264.

Bandeira, O., Rasul, I., 2006. Social networks and technology adoption in northern Mozambique. Econ. J. 116 (514), 869–902.

Barnejee, A., Chandrasekhar, A.G., Duflo, E., Jackson, M.O., 2006. The diffusion of microfinance. Science 341, 1236498.

Barwick, P.J., Liu, Y., Patacchini, E., Wu, Q., 2019. Information, mobile communication, and referral effects. Working Paper 25873, National Bureau of Economic Research.

Bayer, P., McMillan, R., Murphy, A., Timmins, C., 2016. A dynamic model of demand for houses and neighborhoods. Econometrica 84 (3), 893–942.

Bayer, P., Ross, S.L., Topa, G., 2008. Place of work and place of residence: informal hiring networks and labor market outcomes. J. Polit. Econ. 116 (6), 1150–1196.

Beaman, L., Magruder, J., 2012. Who gets the job referral? Evidence from a social networks experiment. Am. Econ. Rev. 102 (7), 3574–3593.

Bhuller, M., Dahl, G.B., Løken, K.V., Mogstad, M., 2020. Incarceration, recidivism, and employment. J. Polit. Econ. 128 (4), 1269–1324.

Billings, S.B., Deming, D.J., Ross, S.L., 2019. Partners in crime. Am. Econ. J.: Appl. Econ. 11 (1), 126–150.

Blanchard, O.J., Katz, L.F., 1992. Regional evolutions. Brook. Pap. Econ. Act. 23 (1), 1–76.

Blondel, V.D., Decuyper, A., Krings, G., 2015. A survey of results on mobile phone datasets analysis. EPJ Data Sci. 4 (10) https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-015-0046-0#citeas.

Blumenstock, J., Chi, G., Tan, X., 2019. Migration and the value of social networks. Discussion Paper 13611, Centre for Economic Policy Research.

Bonacich, P., 1972. Factoring and weighting approaches to clique identification. J. Math. Sociol. 2 (11), 113–120.

Brasche, S., Bischof, W., 2005. Daily time spent indoors in German homes — baseline data for the assessment of indoor exposure of German occupants. Int. J. Hyg. Environ. Health 208 (4), 247–253.

Brown, M., Setren, E., Topa, G., 2016. Do informal referrals lead to better matches? Evidence from a firm's employee referral system. J. Labor Econ. 34 (1), 161–209.

Büchel, K., Ehrlich, M.V., 2021. Cities and the structure of social interactions: Evidence from mobile phone data. Journal of Urban Economics doi:10.1016/j.jue.2020.103276.

Calabrese, F., Smoreda, Z., Blondel, V.D., Ratti, C., 2011. Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. PLoS One 6 (7), e20814.

Calvò-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. Rev. Econ. Stud. 76 (4), 1239–1267.

Calvò-Armengol, A., Zenou, Y., 2012. Social networks and crime decisions: the role of social structure in facilitating delinquent behavior. Int. Econ. Rev. 45 (3), 939–958.

Card, D., 2001. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. J. Labor Econ. 19 (1), 22–64.

Coate, P., Mangum, K., 2019. Fast locations and slowing labor mobility. Working Paper 19–49, Federal Reserve Bank of Philadelphia.

Conley, T.G., Udry, C.R., 2010. Learning about a new technology: Pineapple in Ghana. Am. Econ. Rev. 100 (1), 35–69.

Cosslett, S.R., 1981. Maximum likelihood estimator for choice-based samples. Econometrica 49 (5), 1289–1316.

Costa, D.L., Kahn, M.E., Roudiez, C., Wilson, S., 2018. Persistent social networks: civil war veterans who fought together co-locate in later life. Reg. Sci. Urban Econ. 70, 289–299.

Davis, M.A., Fisher, J.D.M., Veracierto, M., 2016. Gross Migration, Housing and Urban Population Dynamics. Processed, Rutgers University.

Diamond, R., 2016. The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000. Am. Econ. Rev. 106 (3), 479–524.

Domencich, T., McFadden, D.L., 1975. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam.

Dustmann, C., Schönberg, U., Stuhler, J., 2016. The impact of immigration: why do studies reach such different results? J. Econ. Perspect. 30 (4), 31–56.

Eidgenössische Kommunikationskommission, 2015. Tätigkeitsbericht 2015 der Eidgenössische Kommunikationskommission. Eidgenössische Kommunikationskommission (ComCom), Bern.

Finnie, R., 2004. Who moves? A logit model analysis of inter-provincial migration in Canada. Appl. Econ. 36 (16), 1759–1779.

Giulietti, C., Wahba, J., Zenou, Y., 2018. Strong versus weak ties in migration. Eur. Econ. Rev. 104, 111–137.

Glaeser, E.L., 2008. Cities, Agglomeration and Spatial Equilibrium. Oxford University Press, Oxford.

Greenwood, M.J., 1997. Internal migration in developed countries. In: Rosenberg, M.R., Stark, O. (Eds.), Handbook of Population and Family Economics, volume 1B. Elsevier, Amsterdam, pp. 647–720.

Hellerstein, J.K., Kutzbach, M.J., Neumark, D., 2014. Do labor market networks have an important spatial dimension? J. Urban Econ. 79, 39–58.

Ioannides, Y.M., 2013. From Neighborhoods to Nations: The Economics of Social Interactions. Princeton University Press, Princeton.

Jackson, M.O., Rogers, B.W., Zenou, Y., 2020. Networks: an economic perspective. In: Light, R., Moody, J. (Eds.), Oxford Handbook of Social Network Analysis. Oxford University Press, Oxford.

Kelly, M., Ó Gráda, C., 2000. Market contagion: evidence from the panics of 1854 and 1857. Am. Econ. Rev. 90 (5), 1110–1124.

Kennan, J., Walker, J.R., 2011. The effect of expected income on individual migration decisions. Econometrica 79 (1), 211–251.

Kim, J.S., Patacchini, E., Picard, P.M., Zenou, Y., 2020. Spatial Interactions. Processed, Monash University.

Klepeis, N.E., Nelson, W.C., Ott, W.R., Robinson, J.P., Tsang, A.M., Switzer, P., Behar, J.V., Herng, S.C., Engelmann, W.H., 2001. The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants. J. Expo. Anal. Environ. Epidemiol. 11, 211–252.

Kline, P., Moretti, E., 2014. People, places, and public policy: some simple welfare economics of local economic development policies. Annu. Rev. Econom. 6, 629–662.

Koşar, G., Ransom, T., van der Klaauw, W., 2019. Understanding migration aversion using elicited counterfactual choice probabilities. Staff Report 883, Federal Reserve Bank of New York.

MacDonald, J.S., MacDonald, L.D., 1964. Chain migration, ethnic neighborhood formation and social networks. Milbank Mem. Fund Q. 42 (1), 82–97.

Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. Econometrica 45 (8), 1977–1988.

Matz, C.J., Stieb, D.M., Davis, K., Egyed, M., Rose, A., Chou, B., Brion, O., 2014. Effects of age, season, gender and urban-rural status on time-activity: Canadian human activity pattern survey 2 (CHAPS 2). Int. J. Environ. Res. Public Health 11 (2), 2108–2124.

McFadden, D.L., 1973. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York NY, pp. 105–142.

McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. 4 (1), 103–120.
Molloy, R., Smith, C.L., Wozniak, A., 2011. Internal migration in the United States. J. Econ. Perspect. 25 (3), 173–196.
Moneyland, 2018. Switzerland telephone call survey 2018: This is how Swiss make phone calls. Telecom News 19/06/2018, Moneyland.ch AG. https://www.moneyland.ch/en/financial-news/.
Monras, J., 2018. Economic Shocks and Internal Migration. Processed, Universitat Pompeu Fabra.
Moretti, E., 2011. Local labor markets. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, 4. Elsevier, Amsterdam, pp. 1237–1313.
Munshi, K., Rosenzweig, M., 2016. Networks and misallocation: insurance, migration, and the rural-urban wage gap. Am. Econ. Rev. 106 (1), 46–98.
Patacchini, E., Arduini, T., 2016. Residential choices of young Americans. J. Hous. Econ. 34, 69–81.
Roback, J., 1982. Wages, rents, and the quality of life. J. Polit. Econ. 90 (6), 1257–1278.
De la Roca, J., Puga, D., 2017. Learning by working in big cities. Rev. Econ. Stud. 84 (1), 106–142.
Rosen, S., 1979. Wage-based indexes of urban quality of life. In: Miezkowski, P.N., Straszheim, M.R. (Eds.), Current Issues in Urban Economics. Johns Hopkins University Press, Baltimore, MD, pp. 74–104.
Schmidheiny, K., 2006. Income segregation and local progressive taxation: empirical evidence from Switzerland. J. Public Econ. 90 (3), 429–458.
Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M.M., Larsen, J.E., Lehmann, S., 2014. Measuring large-scale social networks with high resolution. PLoS One 9 (4), e95978.
Topa, G., Zenou, Y., 2015. Neighborhood and network effects. In: Duranton, G., Henderson, J.V., Strange, W. (Eds.), Handbook of Regional and Urban Economics, 5. Elsevier, Amsterdam, pp. 561–624.
UNESCOunesco and World Bank, 2018. Culture in City Reconstruction and Recovery. United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris.
Wahba, J., Zenou, Y., 2012. Out of sight, out of mind: migration, entrepreneurship and social capital. Reg. Sci. Urban Econ. 42 (5), 890–903.
Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabasi, A.-L., 2011. Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, pp. 1100–1108.
Zignani, M., Quadri, C., Gaito, S., Rossi, G.P., 2015. Calling, texting, and moving: multidimensional interactions of mobile phone users. Comput. Soc. Netw. 2 (13).