working paper
1909

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

Konstantin Büchel
Maximilian V. Ehrlich
Diego Puga
Elisabet Viladecans-Marsal

March 2019

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 1909
March 2019

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

Abstract

Using anonymised cellphone data, we study the role of social networks in residential mobility decisions. Individuals with few local contacts are more likely to change residence. Movers strongly prefer places with more of their contacts closeby. Contacts matter because proximity to them is itself valuable and increases the enjoyment of attractive locations. They also provide hard-to-find local information and reduce frictions, especially in home-search. Local contacts who left recently or are more central are particularly influential. As people age, proximity to family gains importance relative to friends.

JEL Codes: R23, L14.

Keywords: Social networks, residential mobility.

Konstantin Büchel
University of Bern
konstantin.buechel@vwi.unibe.ch

Maximilian V. Ehrlich
University of Bern
maximilian.vonehrlich@vwi.unibe.ch

Diego Puga
CEMFI
diego.puga@cemfi

Elisabet Viladecans-Marsal
Universitat de Barcelona
eviladecans@ub.edu

Acknowledgement

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 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 Ministerio de Economía y Competitividad (grant ECO2016–75941–R).

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 it. Depending on our residential location choice, there are also substantial differences in the extent to which jobs, education opportunities and amenities are within reach and in who we can interact with. Even when accessible, getting to places and people often requires substantial transit, and close to 10% of our wake time is typically spent traveling, with large 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.[1]

Research on residential location choices tends to focus on determinants that are common across individuals or broad groups, such as job opportunities, housing costs, accessibility, amenities, 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). In practice, migration flows react slowly even in the face of large shocks. Furthermore, gross flows are many times larger than net flows, with apparently similar people simultaneously moving in opposite directions (Davis, Fisher, and Veracierto, 2016, Monras, 2018). 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 study the important role played by the location of each person's social connections in residential location choices. Our emphasis on social networks and how they interact with local characteristics to make preferences across locations more idiosyncratic is consistent with a spatial equilibrium framework. In fact, precisely if we are close to a spatial equilibrium, common determinants will tend to balance out (expensive homes will offset high-paying jobs and lake views), and most moves will be driven by features that are specific to an individual-location pair. 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. Studying the role of local ties on people's attachment to a place also helps us understand why social and family networks can make communities more resilient in the face of economic shocks or natural disasters (UNESCO and World Bank, 2018).

[1] The average person spends at home 15.6 hours per day in the United States, 15.8 hours in Canada, and 15.7 hours 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 minutes per day travelling and 15.2 hours 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, Smith, and Wozniak, 2011), while 5% of cellphone users in our data move across postcodes in Switzerland in a year (table 2 below).

Our analysis shows that taking into account where each person's contacts live doubles our ability to predict who will move and where, helping understand why apparently 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.

Contacts matter for residential location choices for three main reasons. 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. A 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. 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.

Exploring heterogeneity further, we find that people who left recently are particularly influential in determining the subsequent relocation choices of the contacts they had in their previous location. Contacts who are themselves better connected, as measured by their eigenvector centrality in the overall Swiss network, also play a particularly big role. We also examine differences across demographic groups. Interestingly, as people age, proximity to family gains importance relative to friends.

Anonymised cellphone data allows us to combine information about changes in individuals' neighbourhood of residence with the size, spatial distribution and other characteristics of their social network. For this paper, we use information on each individual's social network, derived from anonymised cellphone Call Detail Records (CDRs), in combination with demographic and location attributes. The CDRs correspond to the universe of customers (2.7 million private cellphone lines making 1.8 billion annual calls) of a Swiss telecommunications operator (with a 55% national market share in a country with virtually universal cellphone penetration) over the twelve-month period between June 2015 and May 2016. 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, and 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.

Until recently, few large-scale datasets had information on both network connections and geographical locations. Most data on social networks is collected using survey techniques. As a result, research is often restricted to a few areas and/or focused on developing countries, where data collection is less expensive (e.g. Alatas, Barnejee, Chandrasekha, Hanna, and Olken, 2016). A common alternative approach has been to infer social connections from having two individuals live or work in close proximity or attend the same school (e.g. Bayer, Ross, and Topa, 2008, Billings, Deming, and Ross, 2019). Cellphone CDRs instead provide direct evidence of actual interactions. In addition, unlike networks inferred from co-location, CDRs allow us to relate residential choices to those of contacts living in the same neighbourhood, separately from those of contacts living elsewhere, and separately from those of co-residents with whom there are no interactions. This helps us, for instance, separate the role of contacts in keeping individuals attached to their current neighbourhood from the role they can play in facilitating integration into a new neighbourhood. It also allows us to provide evidence that our results do not merely reflect the sorting of similar people into similar places. We do so by checking that the earlier relocation choices of contacts who used to live in the same neighbourhood matter over and above the choices of strangers with similar characteristics who made the same relocation choices. More recently, research has exploited data from online social networks, such as Twitter, Facebook, or LinkedIn (e.g. Bailey, Cao, Kuchler, and Stroebel, 2018).[2] These data are useful to capture alternative channels to transmit information, but online connections are much more weakly related than calls to direct personal interactions (Stopczynski *et al.*, 2014).

Many economic decisions and outcomes are shaped by networks. See Ioannides (2013) and Jackson, Rogers, and Zenou (2019) for detailed descriptions of this literature.[3] 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 sharing a neighbourhood. However, there are some notable recent exceptions. Perhaps the closest study to ours in terms of topic, even if in a completely different context, is Costa, Kahn, Roudiez, and Wilson (2018). They 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. Other papers, without considering the residential location choice among different neighbourhoods or towns, study how contacts influence rural-urban migration in a developing country context (Munshi and Rosenzweig, 2016, Giulietti, Wahba, and Zenou, 2018, Blumenstock, Chi, and Tan, 2018).[4]

Studying the role of social networks in residential location choices is complicated by several aspects. Individuals may build networks selectively at a place where they are about to move,

[2]They 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.

[3]The topics covered include job market referrals and labour outcomes (e.g. Bayer, Ross, and Topa, 2008, Beaman and Magruder, 2012, Hellerstein, Kutzbach, and Neumark, 2014, Brown, Setren, and Topa, 2016), school performance (e.g. Calvò-Armengol, Patacchini, and Zenou, 2009), technology adoption (e.g. Bandeira and Rasul, 2006, Conley and Udry, 2010, Barnejee, Chandrasekhar, Duflo, and Jackson, 2006), crime and incarceration (e.g. Calvò-Armengol and Zenou, 2012, Bhuller, Dahl, Løken, and Mogstad, 2018, Billings, Deming, and Ross, 2019), nest-leaving (Patacchini and Arduini, 2016), and financial market contagion (Kelly and Ó Gráda, 2000) among others.

[4]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.

creating reverse causality. To address this issue, we explore a pre-determined large-scale network focusing on contacts that the individual already had between six and four months before moving —and increasing the gap between the time window in which we characterise the network and the moving date has no bearing on our results. We also look at the importance of characteristics of contacts that the individual cannot affect —e.g. eigenvector centrality or kinship— and at the importance of second-order links (friends of friends who are not one's friends). Another 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, we begin by including as controls interactions between location characteristics and observable individual characteristics. Since this still leaves open the possibility of sorting on unobservables, we also consider 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. This shows that, while relocations across certain location pairs are particularly common, this is not what drives the effects we measure for social networks. This strategy is possible thanks to the richness of our network data, which due to their comprehensiveness is also much less prone to measurement error than the smaller-scale survey data that are often used to study network effects.

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 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. We also highlight that contacts are an important source of information, and focus on this in section 7. 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 8. Section 9 concludes.

2. Data

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, the use of calls seems particularly appropriate to study how having social connections who live in different places affects the probability of changing residence and moving closer to them. First, because most people use some combination of calls, direct encounters, and text messages to communicate; two people who call each other are very likely to interact more broadly.[5] Second, because calls between users can be measured more reliably on a large scale than direct encounters.[6] Third, because calls are a better indication of close connections and frequent interactions than text messages —particularly between those living far apart.[7]

Data on telephone communications and individual characteristics

The main dataset used in this paper comprises the anonymised Call Detail Records (CDRs) of all calls and text messages 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, a code for the type of communication recorded (call, SMS, or MMS), 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 December 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.

[5]Using CDRs that include information on the transmitting cell tower, Calabrese, Smoreda, Blondel, and Ratti (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 kilometres apart. Similarly, Wang, Pedreschi, Song, Giannotti, and Barabasi (2011) show that the frequency of direct encounters between cellphone users is highly correlated with their frequency of calls.

[6]Direct encounters are usually not observed by researchers but instead inferred from residing or working in close proximity or attending the same school (e.g. Bayer, Ross, and Topa, 2008, Billings, Deming, and Ross, 2019). A more reliable large-scale method involves cross validation using phone calls. For instance, Calabrese, Smoreda, Blondel, and Ratti (2011) proxy face-to-face contact by identifying when two users simultaneously use the same cell tower to call one another in an area where neither lives or works. For small samples, it is possible to measure face-to-face meetings more precisely through other wireless technologies that can detect proximity. Stopczynski *et al.* (2014) issued 1,000 Danish university students voluntarily participating in their study with cellphones and an application that used bluetooth technology to scan for other participants' devices within a 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.

[7]Zignani, Quadri, Gaito, and Rossi (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. Furthermore, 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).

Table 1: Sample representativeness

	Sample	Census	Correlation sample-census at the level of		
			Employment areas	Districts	Municipalities
	(1)	(2)	(3)	(4)	(5)
Individuals	2.1×10^6	6.7×10^6	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.

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.

Table 1 compares cellphone customers in our sample with the overall population of Switzerland. 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). Comparing columns (1) and (2) we see that 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. 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 gets reflected in the age composition of the provider's customer base. Columns (3)–(7) provide 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. We can see that, 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.

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 3,152 potential residential locations, each corresponding to a distinct postcode.[8] We measure location characteristics not just at the postcode level (e.g. housing variables), but within given travel times of this (e.g. the share of the individual's contacts reachable in less than 10 minutes), at the municipal level (2,322 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 house, 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.

Table 2 compares 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. Next 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 minutes of travel time, 32% between postcodes 10–20 minutes apart, 16% between postcodes 20–30 minutes apart, 9% between postcodes 30–40 minutes 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.

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.[9] 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

[8] Our set of 3,152 postcodes excludes a small number of special codes that are not usable for tracking potential residential locations, such as those assigned to large hospitals.

[9] For a discussion on filtering of cellphone data see Blondel, Decuyper, and Krings (2015).

Table 2: 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.

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.

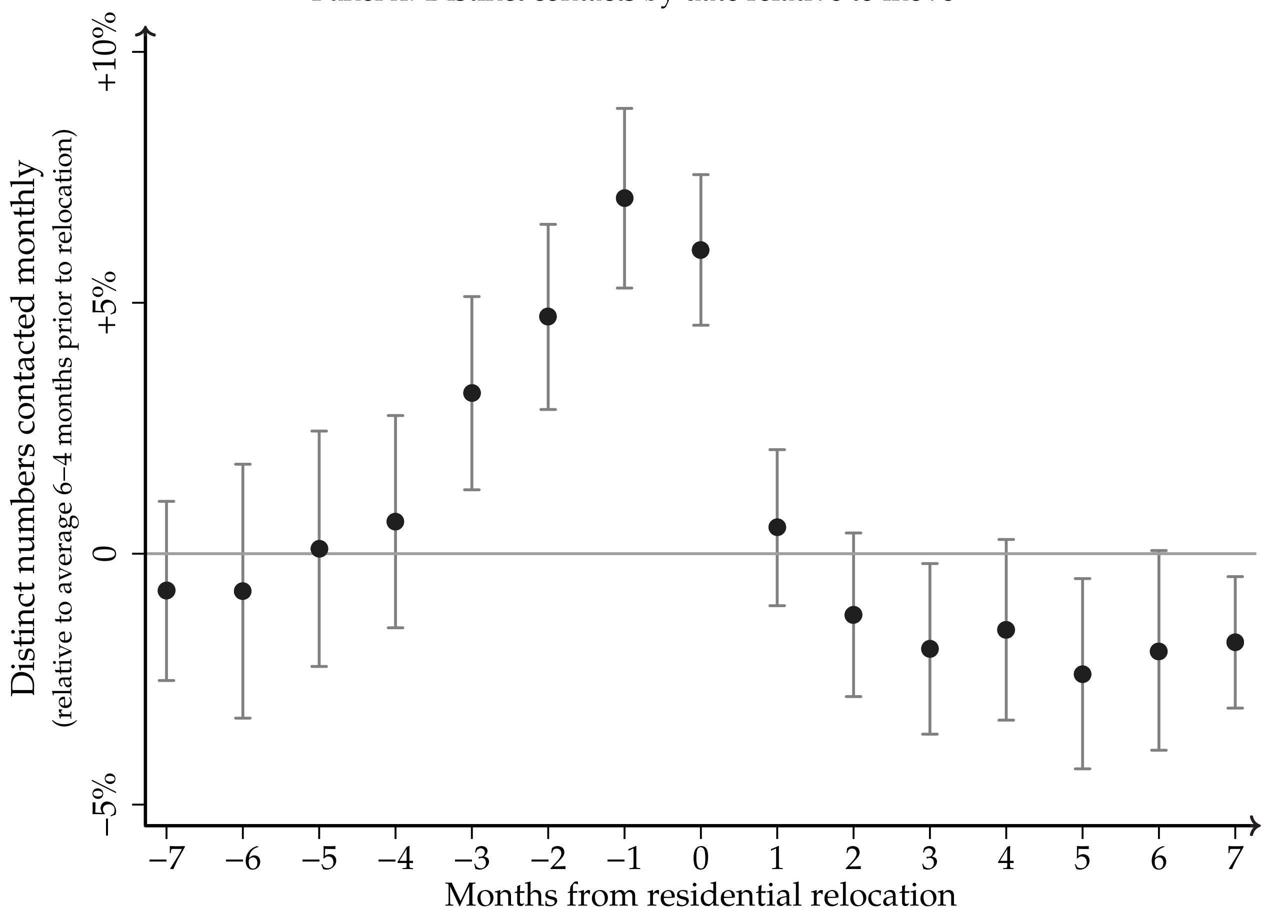
Our starting sample is made up of 2.7 million distinct private cellphone customers making a total of 608 million intra-operator calls to other private cellphone lines. We exclude accidental calls by dropping calls with a duration of less than 10 seconds. We also drop cellphone numbers that display implausibly 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 hours. 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 2,136,093 cellphone customers and 410 million calls.

Defining the social network matrix

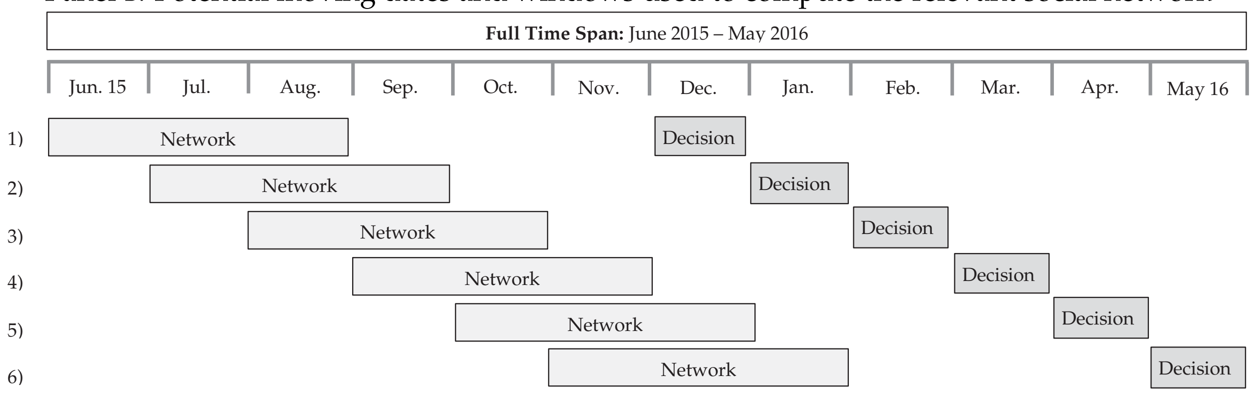
Our primary aim is to study how having social ties living in different places affects the probability of changing residence and moving closer to them. Thus, we would like to characterise each individual's social network at the time of a potential move using only calls that reflect a pre-existing

Figure 1: Timing of relocation decisions

Panel A: Distinct contacts by date relative to move



Panel B: Potential moving dates and windows used to compute the relevant social network



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 figure 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 figure 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 figure 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. In appendix B we show 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. 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 how many contacts individual i has spoken with on the phone during the month 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. Finally, we aggregate postcodes for each individual by travel-time intervals. For this purpose, we construct a distance matrix providing the time it takes to travel by car across any two postcodes in Switzerland. These travel times are obtained from `https://www.search.ch` and correspond to travel by road under normal traffic conditions. 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 minutes travel time of postcode j. Likewise we calculate the share of each individual's contacts who reside within 10–20 minutes, 20–30 minutes, and 30–40 minutes of each postcode. For robustness results shown in appendix B, we also compute the number (instead of share) of each individual's contacts who reside within 0–10, 10–20, 20–30, and 30–40 minutes of each postcode.

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 such characteristics, we focus on elements that may matter more or less depending either on the individual's network of contacts or on their observable demographic characteristics.

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.

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.[10] 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.

[10]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.

3. Framework and estimation

We base our estimation strategy on a two-stage approach. We first analyse the binary decision on *whether* to change residence or not (migration decision). Then, for those who relocate, we study the decision of *where* to relocate (location decision). It is relatively common for papers to consider only the second of these decisions, i.e. the choice of where to locate conditional on moving (e.g. Schmidheiny, 2006). 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.

The indirect utility individual i attains at location j depends on individual characteristics (θ_i), location characteristics (λ_j), and individual-location-specific characteristics (π_{ij}):

$$V_{ij} = \theta_i + \lambda_j + \pi_{ij} + I_{ij,t-1}(\tilde{\lambda}_j + \tilde{\pi}_{ij}) + \epsilon_{ij}, \tag{1}$$

where $I_{ij,t-1}$ is an indicator which is unity if individual i already resided at j in the previous period and zero otherwise. Note that $\tilde{\lambda}_j$, $\tilde{\pi}_{ij}$ capture heterogeneity in the location and individual-location specific components for stayers with $I_{ij,t-1} = 1$ and movers with $I_{ij,t-1} = 0$, while ϵ_{ij} denotes any unobserved preference components. The utility level a given individual derives from a specific location is likely to be different if they were already in the previous period than if they have just moved. This is partly because there are moving costs. It 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 $f(\cdot)$:

$$\pi_{ij} = g(\boldsymbol{N}_{ij}, \boldsymbol{N}_{ij}\check{\boldsymbol{Z}}_j) + f(\boldsymbol{X}_i\boldsymbol{Z}_j). \tag{2}$$

The network component includes a vector $\boldsymbol{N}_{ij}$ which has four columns: the shares of each individual's contacts who reside within 10 minutes, 10–20 minutes, 20–30 minutes, and 30–40 minutes of each potential location. For some specification, we also consider the shares of second-order links within these time intervals and 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 $\boldsymbol{N}_{ij}\check{\boldsymbol{Z}}_j$ isolates the effects of observable location characteristics that make a location j more or less attractive. This informs us about the role of networks in providing information about locations, consumption complementaries, and about how networks help alleviating search frictions. To this end our selection of local characteristics aims at characteristics that are difficult to observe from elsewhere, or reflect search frictions. In particular the columns of $\check{\boldsymbol{Z}}_j$ include measures for local housing market, cultural amenities, local crime rates, and availability of childcare. Since these observable amenities and disamenities are only a subset of relevant local factors we also estimate more generic specifications where the location fixed effect (λ_j) is interacted with the network measures.

The other individual-location specific component $f(\cdot)$ captures interactions between location characteristics ($\boldsymbol{Z}_j$) and individual, non-network characteristics ($\boldsymbol{X}_i$). The location characteristics include factors that differ in their relevance for different groups of the population such as the availability of childcare as well as information about local majority language, average age, average household size, share of homeowners, share of foreign immigrants, population density, employment areas, and income taxes. The individual specific observables in $\boldsymbol{X}_i$ include: age, gender, language of correspondence, previous location, and an indicator for the individual having multiple cellphone numbers. These factors absorb sorting of heterogeneous individual to different places to some degree, but —as usual— this only partly accounts for sorting to the extent that individual characteristics are observable.

Estimating the probability of changing residential location. According to equation (2), an individual will decide to migrate if there exists a location j which provides higher utility than the current residence r:

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

Put differently, an individual decides to migrate if the utility at the current place of residence r drops below an individual-specific utility threshold. We estimate the probability of moving by using linear probability as well as logit models. As we pool six moving windows (depicted in figure 1), we include time fixed effects in all empirical specifications of (3).

Estimating the residential location choice. The second-stage of our approach explores the location choice of those individuals that decided to move. Conditional on $j \neq r$ we estimate the likelihood that a location alternative j provides the highest level of utility among the movers' choice set of locations:

$$\text{Prob}[\max_{k \neq r}(V_k)] = V_{ij}, \tag{4}$$

where utility is composed of location- and individual-location specific factors as in (1). Assuming that ϵ_{ij} is drawn from an extreme value distribution, the probability that i chooses location j is

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

which can be estimated using a conditional logit model. As in the migration decision model, we include time-fixed effects in all specifications.

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 $j = r$. In addition, the latter would assume a common cost of moving for all locations which according to our findings does not hold true in the data. We find a low correlation between the location fixed-effects of both stages which invalidates a one-stage approach. Finally, the two-stage approach allows for heterogeneous effects of individual-location interactions, network and network-location interactions in the migration and location decisions. This seems quite relevant as one may expect that for instance that the scarcity of house vacancies interacted with the strength of the network at a moving destination matters for location choice as friends may help to find a home whereas it seem unlikely to be relevant once already settled i.e. in the decision of whether to move away.

Choice set in the location decision. In principle, all 3,152 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 is one if the individual has chosen location alternative k and zero otherwise. The linear model is estimated on the full set of potential location alternatives as well as for restricted choice sets. Second, we explore reasonable restrictions of the choice set based on network information. It turns out that 95.3% of movers relocate to a postcode where they have at least one pre-existing contact within 40 minutes. On average, an individual has about 500 alternatives with at least one contact within 40 minutes. Based on this, we use all location alternatives with a contact within a radius of 40 minutes and sample the remaining ones where we attach observed, individual-specific, weights for the two groups of sampled and non-sampled choice alternatives.[11] Finally, we also perform our estimations considering only location alternatives with a contact within a radius of 40 minutes. As we show below, all of these choice sets yield 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 (UNESCO and World Bank, 2018). In an internal migration context, Coate and Mangum (2018) argue that most of the recent decline in mobility within the United States is due to increasing 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 3, we estimate the probability of changing residence to a different postcode as a function of individual, location, and individual-location characteristics for 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 measure is the share of the person's contacts residing within a certain travel time from their current home. In addition, we also consider the share of second-order links (i.e. friends of friends who are not one's friends)

[11] Since we observe the true number of location alternatives with and without contact within a radius of 40 minutes we can use a stratified sample and adjust the estimates with the respective sampling weights (Manski and Lerman, 1977, Cosslett, 1981).

Table 3: Probability of changing residential location

Dep. var.: Probability of changing residential location	Linear prob. model			Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of contacts						
0–10 min	-3.008*** (0.039)	-3.596*** (0.050)	-3.199*** (0.061)	-1.426*** (0.019)	-1.606*** (0.022)	-1.518*** (0.029)
10–20 min		-1.450*** (0.059)	-1.327*** (0.060)		-0.478*** (0.023)	-0.458*** (0.024)
20–30 min		-0.717*** (0.072)	-0.671*** (0.072)		-0.211*** (0.028)	-0.204*** (0.028)
30–40 min		-0.365*** (0.086)	-0.343*** (0.086)		-0.151** (0.034)	-0.148*** (0.034)
Share of 2nd-order contacts 0–10 min			-1.025*** (0.094)			-0.217*** (0.049)
Degree centrality (# contacts)	0.010*** (0.001)	0.011*** (0.001)	0.011*** (0.001)	0.005*** (0.000)	0.005*** (0.000)	0.005*** (0.000)
Long-term resident	-0.815*** (0.022)	-0.798*** (0.022)	-0.795*** (0.022)	-0.327*** (0.009)	-0.321*** (0.009)	-0.321*** (0.009)
Speaks same language as majority	-0.508*** (0.076)	-0.388*** (0.077)	-0.379*** (0.077)	-0.123*** (0.027)	-0.083** (0.027)	-0.081* (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
Avg. probability	4.9%	4.9%	4.8%	4.6%	4.7%	4.9%
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). Avg. probability refers to the predicted probability of moving between December 2015 and May 2016 for a male individual in the 25–34 age bracket, who speaks German like the local majority, is not a long-term resident, has the sample mean values for their contact network. 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.

located within a certain travel time from their current residence. The main individual measure is the person's degree centrality (i.e. the total number of contacts the individual has). As explained in section 2 and illustrated in panel B of figure 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. Later, we discuss the implications of also incorporating into the analysis the frequency of calls with different contacts.

In addition to degree centrality, we control for other observable individual characteristics (language, age, and gender). Location characteristics are captured by a location fixed effect. As additional individual-location characteristics, we include an indicator for whether the individual shares the local majority language and 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. 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 minutes 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)[12] 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 $4.8 - 0.289 \times (-3.008)$).

Degree centrality, defined as the individual's total number of contacts, enters positively and statistically significantly. 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 those who are long-term residents 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 minutes, 20–30 minutes, and 30–40 minutes of the current postcode, we see that they are all negative and statistically significant. In terms of magnitude, the deterrent effect on mobility of having a larger share of contacts within 30–40 minutes of the current residence (instead of the baseline over 40 minutes away) is about one-half as large as the deterrent effect of having them within 20–30 minutes of the current residence. In turn the effect of having a larger share of contacts within 20–30 minutes is about one-half as large of the effect of having them within 10–20 minutes, which in turn is about one-half as large of the effect of having them within 0–10 minutes.

To the extent that contacts are a key channel for acquiring information, a strong concentration of social ties close to the current residence will lead to greater asymmetry in the cost of getting information about amenities, education and job opportunities, etc. between the current and alternative locations. The share of the individual's contacts located within a given travel time from their current residence combines both the direct enjoyment of interacting with contacts and the information advantage. In column (3), we also add the share of the individual's second-order contacts located within 10 minutes 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. Thus, this variable isolates a pure information effect. If the friends of a person's friends are more concentrated locally they can provide more information (about a trendy new restaurant, a job opportunity, etc.) that would not be as easy to get elsewhere. There is not a direct benefit of interactions because what distinguishes second-order from first-order links is that for the former

[12] See appendix A 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 A.1.

direct interactions have not taken place. We see that the share of second-order links located within 0-10 minutes from the individual's current location also makes a change of residence less likely. This is clear evidence that networks matter greatly for information gathering and that information is an important determinant of residential location choices. At the same time, note that we cannot interpret the difference between the coefficients on first-order and second-order links within 10 minutes 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).

The results for the logistic model of columns (3) and (4) 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 minutes 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.

5. The residential location choice

We now turn from the decision to change residential location to the decision of where specifically to go. Table 4 gives the results for the second step of our analysis, where we study, conditional on deciding to move, the role of a person's social network in choosing their new residential location. The table is now estimated using only movers and identifies coefficients based solely on variation within a given individual across locations. We have 47,214 individuals who move to a different postcode in our data 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 network of contacts with whom the individual spoke on their cellphone six to four months before the moving date.

The dependent variable in table 4 is the probability that an individual mover chooses a specific location over all other alternatives, depending on location and individual-location characteristics. As before, residential locations are defined as each of the over 3,000 postcodes in Switzerland. Estimation with such a large set is obviously 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 minutes. Focusing on such locations reduces the choice set for the average individual to about 540 postcodes. In table 4, we provide results estimated using this restricted set of locations for each individual. Nevertheless, in table B.3 in the appendix we show that the results obtained in table 4 remain nearly identical if we also randomly sample 100 locations from the set where the individual has no contacts within 40 minutes (and attach the relevant individual-specific weights to these sampled alternatives). The same table also shows that results remain essentially unchanged even if we use the full set of location alternatives.[13]

[13]For table 4, considering all locations merely increases computing time. For the more demanding estimations further below, estimation considering all 3,152 postcodes as relevant alternatives for every individual becomes computationally unfeasible. It is for this reason that 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 minutes.

Table 4: Residential location choice

Dep. var.: Prob. of choosing a location conditional on moving	Linear prob. model			Conditional logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of contacts						
0–10 min	9.257*** (0.015)	8.855*** (0.015)	9.085*** (0.015)	8.825*** (0.035)	8.197*** (0.036)	6.263*** (0.027)
10–20 min	0.948*** (0.007)	0.909*** (0.007)		5.088*** (0.033)	4.780*** (0.033)	
20–30 min	0.049*** (0.005)	0.045*** (0.005)		2.457*** (0.035)	2.323*** (0.035)	
30–40 min	-0.037*** (0.004)	-0.038*** (0.004)		0.567*** (0.040)	0.497 (0.040)	
Within same employment area	0.109*** (0.002)	0.105*** (0.002)	0.177*** (0.002)	0.936*** (0.015)	0.964*** (0.015)	1.646*** (0.013)
Within preferred language region	0.012*** (0.003)	0.010** (0.003)	0.029*** (0.003)	0.543*** (0.030)	0.599*** (0.030)	0.875*** (0.028)
Return migration		15.318*** (0.036)	15.386*** (0.036)		2.415*** (0.025)	2.765*** (0.026)
Individual x location controls	No	Yes	Yes	No	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.023	0.030	0.030	-	-	-
Pseudo R^2	-	-	-	0.231	0.253	0.207
N	25,555,189	25,555,189	25,555,189	25,342,595	25,342,595	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)-(3) and at the employment region level in columns (4)-(6). Individual $\times$ location controls are 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. Return migration indicates the individual was a resident at the same location at an earlier time in 2012–2015. The pseudo R^2 in columns (4)-(6) is calculated followingMcFadden (1973). ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

Throughout our regressions analysing residential location choices, instead of considering specific location characteristics (e.g. housing prices, local tax rates, 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. We consider specific location characteristics further below, to the extent that we think these may matter more or less depending on each individual's network.

Columns (1) to (3) estimate the probability of choosing a specific new residential location using a linear probability model, while columns (4) to (6) do so using a conditional logit. We begin with the linear probability model and will focus on this for much of the analysis because it has two important advantages over the conditional logit. Most importantly, because in the linear probability model it is computationally feasible to include a location fixed effects for each of the over 3,000 postcodes in Switzerland, whereas in the conditional logit we can only include location fixed effects for each of the country's 16 employment regions. 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.

The results in column (1) indicate that the prior presence of local contacts increases the likelihood of choosing a location alternative. The estimated coefficient for the share of contacts within 10 minutes travel time is positive and significant. It indicates that having 10% more local contacts within this distance increases the base probability of choosing that location relative to all others by almost one percentage point (0.10×9.257) —a substantial effect given that there are over 3,000 postcodes to choose from and that the average individual has pre-existing contacts within 40 minutes 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 located within 0–10 minutes 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 a further 10 minutes away only matter about one-tenth as much for the decision of where to relocate. When looking at the decision of whether to move (instead of where to move) in table 3, we saw that in that context contacts located a further 10 minutes away mattered about one-half as much.

The results also show that choosing a given new postcode is much 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.

Relative to column (1), column (2) incorporates a large series of interactions between individual characteristics and location characteristics, mostly meant to capture how well they match.[14] These control 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). We see that adding these individual $\times$ location controls increases the explanatory power of our model, but does not diminish the importance of the variables characterising where contacts are located. In the following section, we develop further strategies to deal with sorting.

As discussed in the introduction, much of the literature is limited by the available data to assigning a special role only to the individual's birth location or to locations where the individual has lived before. One may worry that, as sophisticated and precise as they may be, 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. To show that this is not the case, in column (2) we add an indicator for return migration, which takes value one if the individual was a resident at the same location at an earlier time. The corresponding coefficient

[14]We include an interaction between an indicator for the individual's preferred language of correspondence 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, in addition to the return migration indicator discussed separately.

Table 5: Predictive power

	Location FE & indiv. x location controls, & share local contacts	Location FE, indiv. x location controls	Location FE, indiv. x location controls, & # local contacts
	(1)	(2)	(3)
Correct predictions at postcode level	10.2%	5.1%	9.7%
Correct predictions at district level	43.8%	18.2%	42.0%
R^2 linear probability model	0.030	0.015	0.028
Pseudo R^2 conditional logit	0.253	0.116	0.243

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 4 column (3). Column (2) corresponds to the same estimation as column (1) without the share of contacts 0–10 min variable. Column (3) corresponds to the same estimation as column (1) but using number of contacts 0–10 min instead of share of contacts 0–10 min.

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 minutes. We will treat this as our base specification for subsequent tables. Columns (4) to (6) replicate the estimations of column (1) to (3) using a conditional logit instead of a linear probability model, finding comparable results.

To get a better idea of the extent to which taking into account where each person's contacts live helps us understand their residential location decisions, in table 5 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 the estimation of table 4 column (3). This specification can guess the exact postcode to where 10.2% of movers relocate. Column (2) corresponds to the exact same specification, removing only the share of contacts within 0-10 minutes travel time, and this makes the percentage of correct predictions at the postcode level drop to 5.1%.

Since guessing the exact postcode chosen may be excessively demanding, in the second row of table 5 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. Once again, when we include share of local contacts, the percentage of correctly predicted relocations is over twice as large, 43.8% compared 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 doubles our ability to predict where individuals relocate.

Column (3) of table 5 reports the predictive power of the same specification of column (1), but using the number (instead of the share) of the individual's contacts within 10 minutes travel distance. The predictive power of this alternative specification is still high, but below that of the specification using the share. Tables B.1 and B.2 in appendix B replicate tables 3 and 4 using the number (instead of the share) of the individual's contacts within 10 minutes travel distance. The distribution of each individual's contacts in terms of numbers and in terms of shares are very highly correlated, so we must use one or the other but not both. Given that using the share provides more accurate prediction and slightly better fit, we keep focusing on shares.

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, Schönberg, and Stuhler, 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: 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

Table 6: Chain mobility

Dep. var.: Prob. of choosing a location conditional on moving	Linear prob. model			Conditional logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of contacts 0–10 min	9.085*** (0.015)	8.518*** (0.015)	8.166*** (0.017)	6.263*** (0.027)	5.930*** (0.027)	5.568*** (0.030)
& recent movers			-0.459*** (0.051)			1.207*** (0.105)
& from same origin			9.107*** (0.094)			3.821*** (0.191)
Share non-contact movers from same origin		88.881*** (0.470)	88.160*** (0.470)		40.529*** (0.795)	40.276*** (0.794)
Within same employment area	0.177*** (0.002)	0.119*** (0.002)	0.117*** (0.002)	1.646*** (0.013)	1.574*** (0.013)	1.567*** (0.013)
Within preferred language region	0.029*** (0.003)	0.014*** (0.003)	0.015*** (0.003)	0.875*** (0.028)	0.858*** (0.029)	0.859*** (0.029)
Return migration	15.386*** (0.036)	15.275*** (0.036)	15.261*** (0.036)	2.765*** (0.026)	2.739*** (0.026)	2.739*** (0.027)
Individual × location controls	Yes	Yes	Yes	Yes	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.030	0.031	0.031			
Pseudo R^2				0.207	0.211	0.213
N	25,555,189	25,555,189	25,555,189	25,342,595	25,342,595	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)-(3) and at the employment region level in columns (4)-(6). Recent movers captures the additional effect of contacts who, in addition to residing in any postcodes that can be reached by car within 10 minutes, moved there between January 2013 and three months prior. From same origin further restricts these to those who moved from within 10 minutes 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 minutes 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 minutes. Individual × location controls as in table 4. The pseudo R^2 in columns (4)-(6) is calculated following McFadden (1973). ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

with location characteristics as controls to account for sorting on observables, as we have done in section 5. However, the possibility of sorting on unobservables being important remains.

In table 6, we take advantage of the richness of our data to better account for sorting and to provide more direct evidence of chain mobility. Column (1) in table 6 corresponds to our benchmark specification in column (3) of table 4, where we include the share of the individual's contacts within 10 minutes travel time of a potential destination postcode as a key determinant of the probability of relocating to that postcode. We are worried that the results that potential new locations with more pre-existing contacts nearby are more likely to be chosen may partly reflect a tendency of similar individuals (more likely to be friends) to relocate across the same postcodes more generally, irrespective of network effects. Column (2) accounts for this possibility by 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 minutes 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 minutes 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 minutes 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 6 we further disentangle the effect of local contacts on relocation decisions. Specifically, we allow these local contacts to matter differently if they arrived at the potential new location recently than if they have been there for some time, and also if they moved from the same original destination than if they moved from a different one. 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 minutes 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 minutes, 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 minutes 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. Once gain, the conditional logit specifications give a similar message, except perhaps for the usefulness of recent movers from a different location.

7. The role of information

Our results so far show that individuals are more likely to leave their current location if they have few contacts nearby and if their contacts are themselves not well connected locally. In choosing where to go, they are also much more likely to pick a neighbourhood where they already know people. A key reason why already knowing people in a prospective new neighbourhood matters so much is that they can provide useful information. In the context of choosing a residential location, there are three aspects to information gathering. One aspect is having sufficient information prior to the move to be able to rank a prospective neighbourhood above others. 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, for instance because childcare is readily available) are more difficult to observe from far away, and this creates an informational asymmetry between areas where one knows people who are likely to have and transmit this information and areas where one knows no-one who can help. A second aspect to information gathering concerns advice that one may wish to gather through the social network on a regular basis after moving. For instance, a neighbourhood may feature a variety of cultural events on a regular basis or have a trendy nightlife scene, but to fully take advantage of this it is useful to know other locals who can share tips of where to go or who may even join in. A third aspect to information gathering concerns searching in markets subject to frictions. In the Swiss context, a country with virtually full employment but with extremely tight housing markets, these friction are most relevant when looking for a new home. Given that houses and apartments for rent or purchase are often gone 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.

To show that all three aspects of information gathering matter, in columns (1) and (3) of table 7 we add to our to our benchmark specification (columns (3) and (6) of table 4, using a linear probability model and a conditional logit respectively) interactions between our main variable measuring the spatial distribution of the individual's social network and relevant measures of local characteristics. 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 minutes and the childcare slots to pupil ratio is positive and significant. Note that the childcare slots to pupil ratio does not appear uninteracted in the regression because we include postcode fixed effects that will absorb this and any other local characteristics.

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 minutes and

Table 7: Interactions between contacts and local characteristics

Dep. var.: Prob. of choosing a location conditional on moving						
	Cond. logit	Linear prob. model				
	(1)	(2)	(3)	(4)	(5)	(6)
Share of contacts 0–10 min	8.255*** (0.040)	9.038*** (0.019)	6.620*** (0.016)	6.394*** (0.017)	7.198*** (0.019)	6.887*** (0.019)
& central				2.992*** (0.052)		
& strong (duration)					-1.610*** (0.038)	
& strong (frequency)						-0.570*** (0.039)
Share of contacts 0–10 min × childcare slots	0.670* (0.299)	6.724*** (0.243)				
× recent crimes	-0.303*** (0.020)	-1.603*** (0.014)				
× cultural events	0.003*** (0.001)	0.042*** (0.001)				
× housing turnover	-18.479*** (2.332)	-238.758*** (1.623)				
× location fixed effect			13.601*** (0.032)	12.239*** (0.034)	14.577*** (0.039)	14.265*** (0.039)
Share of contacts 0–10 min & central × loc. fixed effect				6.655*** (0.115)		
& strong (duration) × loc. f. e.					-5.967*** (0.081)	
& strong (frequency) × loc. f. e.						-4.836*** (0.081)
Share non-contact movers from same origin	17.641*** (0.779)	78.734*** (0.478)	80.333*** (0.475)	88.199*** (0.469)	88.142*** (0.469)	89.057*** (0.469)
Individual × location controls	Yes	Yes	Yes	Yes	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	–	0.033	0.038	0.038	0.038	0.037
Pseudo R^2	0.275	–	–	–	–	–
N	25,313,251	25,538,167	25,555,189	25,555,189	25,555,189	25,555,189

Notes: Dependent variable is expressed as a percentage in the linear probability model. 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 in all columns except (3). In column (3) we estimate location fixed effects at the employment region level while controlling for the estimated fixed effect at the postcode level from column (2) to capture within-employment-region variation. 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 minutes 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 minutes. Individual × location controls includes, in addition to the reported variables, all those in table 4. The pseudo R^2 in column (3) is calculated following McFadden (1973). ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

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 minutes 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.

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 minutes 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.

In column (3) of table 7, instead of measuring whether local contacts matter in combination with individual local characteristics, we summarise all such characteristics with a postcode fixed effect centred at zero.The positive and significant interaction term between the share of contacts living within 10 minutes and the postcode fixed effect indicates that postcodes that are particularly attractive due to their full set of characteristics (those with a large positive fixed effect) are even more likely to be chosen by people who have more of their contacts living nearby, either because they let them know about how nice they are or because sharing local amenities with friends and family makes them even more enjoyable. In table B.4 in the appendix we show that these results remain nearly identical if we use the full set of location alternatives or different random samples of location alternatives where the individual has no contacts within 40 minutes.

The quality and the amount of information that an individual can obtain from their social network depends on the quality of the contacts in the network (more or less central) and also on the intensity of the ties (stronger or weaker) (see Ioannides, 2013, Giulietti, Wahba, and Zenou, 2018). In columns (4) to (6) of table 7 we re-estimate the specification of column (2), now exploring the position of contacts in the network and the intensity of the links. In column (4) we add to the specification of column (3) "share of contacts 0–10 min & central," 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 minutes, are in the top 10% in terms of eigenvector centrality in the overall 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 particularly central local contacts over and above that of the average local contact. We see that central contacts are particularly important drivers of location choices. The interaction between "share of contacts 0–10 min & central" and location fixed effects suggests that central contacts are specially useful as centres of information transmission, which is natural given that centrality measures precisely being more connected to people who themselves are particularly connected.

Columns (5) and (6) consider instead two measures of how strong is the link between the mover and each of their contacts, where strength is measured by the frequency of calls in column (6)

Table 8: Age relevance of amenities

Dep. var.: Prob. of choosing a location conditional on moving	Child Care		Events	
	Age 25–44	Age $\geqslant$ 45	Age 15–34	Age $\geqslant$ 35
	(1)	(2)	(3)	(4)
Share of contacts				
0–10 min	8.390*** (0.018)	7.446*** (0.035)	8.447*** (0.019)	7.323*** (0.025)
Share of contacts within 10 min				
$\times$ childcare slots	8.690*** (0.228)	0.646 (0.425)		
$\times$ cultural events (young target audience)			0.014*** (0.001)	-0.032*** (0.002)
$\times$ cultural events (broad target audience)			0.265*** (0.003)	0.165*** (0.003)
Share non-contact movers from same origin	79.476*** (0.537)	89.454*** (1.046)	79.244*** (0.585)	85.284*** (0.821)
Individual $\times$ location controls	Yes	Yes	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes
R^2	0.033	0.028	0.036	0.026
N	20,596,304	4,941,863	17,129,669	8,425,520

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 $\times$ location controls as in table 4. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

and by the combined duration of calls in column (7). We see that weaker links are an important determinant of migration decisions. Our finding that movers tend to follow people they know who migrated recently from the same origin (table 6) but that weak links are also relevant (table 7) is in line with the results of Giulietti, Wahba, and Zenou (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 important for information gathering.

Table 8 explores the role of contacts in transmitting relevant information further, by focusing on amenities that may be more or less relevant depending on demographics. In columns (1) and (2) we revisit the interaction between the share of contacts living within 10 minutes and the childcare slots to pupil ratio. 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, we see 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. In a similar vein, in columns (3) and (4) we separate local cultural events into those likely to appeal to a young audience and those likely to appeal to a broad target audience. 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 is a DJ session, while older people learn

Table 9: Information via second-order links

Dep. var.: Prob. of choosing a location conditional on moving	(1)	(2)	(3)
Share of contacts 0–10 min	6.012*** (0.019)	5.083*** (0.009)	6.544*** (0.024)
Share of 2nd-order contacts 0–10 min	6.925*** (0.035)	4.420*** (0.037)	7.388*** (0.044)
Share of contacts within 10 min			
× childcare slots			3.703*** (0.325)
× recent crimes			-1.067*** (0.019)
× cultural events			0.022*** (0.001)
× housing turnover			-145.088*** (2.191)
× location fixed effect		6.949*** (0.091)	
Share of 2nd-order contacts within 10 min			
× childcare slots			6.605*** (0.575)
× recent crimes			-1.475*** (0.033)
× cultural events			0.049*** (0.002)
× housing turnover			-225.827*** (3.853)
× location fixed effect		8.991*** (0.066)	
Share non-contact movers from same origin	73.919*** (0.468)	74.755*** (0.469)	71.938*** (0.478)
Individual × location controls	Yes	Yes	Yes
Location fixed effects	Yes	Yes	Yes
R^2	0.033	0.039	0.034
N	25,555,189	25,555,189	25,538,167

Notes: All columns estimated using a linear probability model, with the dependent variable expressed as a percentage. Second-order links exclude first-order links. Location fixed effects defined at the postcode level. Individual × location controls as in table 4. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

from their contacts about how unpleasantly noisy this was).

After showing evidence that contacts are an important source of useful information and that this matters for location choices, we now explore if information coming indirectly from friends of friends is also relevant. In column (1) of table 9 we add to our benchmark specification (column (3) of table 4) the share of the individual's second-order contacts located within 10 minutes 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. One may ask a friend about available houses in heir neighbourhood and they may not know of a suitable one but can ask their friends and come back with suggestions. We can see that the coefficients on first-order and second-order links are almost the same. This points to the great importance of indirect information gathering for relocation decisions. In column (2) we reinforce this message by interacting both the share

Table 10: Relative importance of friends and family by age

	All		Age 25–34		Age 35–54		Age $\geq$ 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 $\times$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 $\times$ 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)
Individual $\times$ 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 $\times$ location controls as in table 4. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

of first-order links and the share of second-order links with a postcode fixed effect. As before, this fixed effect measures the extent to which a postcode is particularly attractive due to their full set of characteristics. Both interaction terms are positive and significant. Once we include the interactions with the postcode fixed effect, the coefficient on the (uninteracted) local share of contacts is smaller for second order links while the coefficient on the interaction with the postcode fixed effect is larger for second order links. This suggests that friends of friends have a particularly strong role for information gathering whereas direct friends also have an important role due to the enjoyment of direct interactions with them. In column (3) we include interactions of first and second order links with the observable location characteristics instead of the fixed effects. These results confirm the role of second order links for information gathering as the coefficients are always significant and show the expected sign. Note however, that the number of second order links is far higher than the number of first order links (by a factor of 16, see table A.1 in the appendix) which implies that the marginal effect of an additional second order link is smaller than the one of a first order link.

8. 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.

Table 10 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 minutes 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 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.

9. 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 locations where to live. We do so using network characteristics obtained from anonymized cellphone data that reflect actual interactions, in combination with demographic and location attributes. Our estimation strategy is organised 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 minutes 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 minutes having an effect at least an order of magnitude greater than the rest.

Already knowing people in a prospective new neighbourhood matters so much partly because they can provide useful information. In the context of choosing a residential location, there are three aspects to information gathering. One aspect is having sufficient information prior to the move to be able to 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 on a regular basis after moving. For instance, a neighbourhood may feature a variety of cultural events on a regular basis or have a trendy nightlife scene, but 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 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. Friends of friends have a particularly strong role for information gathering and weaker direct links also matter on this respect. Direct friends, in addition to providing information and reducing frictions, also have an important role due to the enjoyment of direct interactions with them. Movers tend to follow people they know who migrated recently from the same origin and can help them settle at the new location. This does not merely reflect a tendency of similar individuals (more likely to be friends) to relocate across the same postcodes more generally: known movers from the same origin to a given location continue to matter similarly if we control for non-contact movers from the same origin to a given location. When distinguishing between the influence of friends and family, we find that both mater but proximity to family gains importance with age.

References

Alatas, Vivi, Abhijit Barnejee, Arun G. Chandrasekha, Rema Hanna, and Benjamin A. Olken. 2016. Network structure and the aggregation of information: Theory and evidence from Indonesia. *American Economic Review* 106(7): 1663–1704.

Altonji, Joseph G. and David Card. 1991. The effects of immigration on the labor market outcomes of less-skilled natives. In John M. Abowd and Richard B. Freeman (eds.) *Immigration, Trade and the Labor Market*. Chicago, IL: Chicago University Press, 201–234.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy* 126(6): 2224–2276.

Bandeira, Oriana and Imran Rasul. 2006. Social networks and technology adoption in northern Mozambique. *Economic Journal* 116(514): 869–902.

Barnejee, Abhjit, Arub G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2006. The diffusion of microfinance. *Science* 341: 1236498.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy* 116(6): 1150–1196.

Beaman, Lori and Jeremy Magruder. 2012. Who gets the job referral? evidence from a social networks experiment. *American Economic Review* 102(7): 3574–3593.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2018. Incarceration spillovers in criminal and family networks. Working Paper 24878, National Bureau of Economic Research.

Billings, Stephen B., David J. Deming, and Stephen L. Ross. 2019. Partners in crime. *American Economic Journal: Applied Economics* 11(1): 126–50.

Blanchard, Olivier Jean and Lawrence F. Katz. 1992. Regional evolutions. *Brooking Papers on Economic Activity* 23(1): 1–76.

Blondel, Vincent D., Adeline Decuyper, and Gautier Krings. 2015. A survey of results on mobile phone datasets analysis. *EPJ Data Science* 4(10).

Blumenstock, Joshua, Guanghua Chi, and Xu Tan. 2018. Migration and the value of social networks. Processed, University of California, Berkeley.

Bonacich, Phillip. 1972. Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology* 2(11): 113–120.

Brasche, Sabine and Wolfgang Bischof. 2005. Daily time spent indoors in German homes — baseline data for the assessment of indoor exposure of German occupants. *International Journal of Hygiene and Environmental Health* 208(4): 247–253.

Brown, Meta, Elizabeth Setren, and Giorgio Topa. 2016. Do informal referrals lead to better matches? Evidence from a firm's employee referral system. *Journal of Labor Economics* 34(1): 161–209.

Calabrese, Francesco, Zbigniew Smoreda, Vincent D. Blondel, and Carlo Ratti. 2011. Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLoS ONE* 6(7).

Calvò-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2009. Peer effects and social networks in education. *The Review of Economic Studies* 76(4): 1239–1267.

Calvò-Armengol, Antoni and Yves Zenou. 2012. Social networks and crime decisions: The role of social structure in facilitating delinquent behavior. *International Economic Review* 45(3): 939–958.

Card, David. 2001. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics* 19(1): 22–64.

Coate, Patrick and Kyle Mangum. 2018. Fast locations and slowing labor mobility. Processed, American Institute for Economic Research;.

Conley, Timothy G. and Christopher R. Udry. 2010. Learning about a new technology: Pineapple in Ghana. *American Economic Review* 100(1): 35–69.

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

Costa, Dora L., Matthew E. Kahn, Christopher Roudiez, and Sven Wilson. 2018. Persistent social networks: Civil war veterans who fought together co-locate in later life. *Regional Science and Urban Economics* 70: 289–299.

Davis, Morris A., Jonas D.M. Fisher, and Marcelo Veracierto. 2016. Gross migration, housing and urban population dynamics. Processed, Rutgers University.

Diamond, Rebecca. 2016. The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000. *American Economic Review* 106(3): 479–524.

Domencich, Tom and Daniel L. McFadden. 1975. *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North-Holland.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler. 2016. The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives* 30(4): 31–56.

Eidgenössische Kommunikationskommission. 2015. *Tätigkeitsbericht 2015 der Eidgenössische Kommunikationskommission*. Bern: Eidgenössische Kommunikationskommission (ComCom).

Giulietti, Corrado, Jackline Wahba, and Yves Zenou. 2018. Strong versus weak ties in migration. *European Economic Review* 104: 111–137.

Glaeser, Edward L. 2008. *Cities, Agglomeration and Spatial Equilibrium*. Oxford: Oxford University Press.

Hellerstein, Judith K., Mark J. Kutzbach, and David Neumark. 2014. Do labor market networks have an important spatial dimension? *Journal of Urban Economics* 79: 39–58.

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

Jackson, Matthew O., Brian W. Rogers, and Yves Zenou. 2019. Networks: An economic perspective. In Ryan Light and James Moody (eds.) *Oxford Handbook of Social Network Analysis*. Oxford: Oxford University Press.

Kelly, Morgan and Cormac Ó Gráda. 2000. Market contagion: Evidence from the panics of 1854 and 1857. *The American Economic Review* 90(5): 1110–1124.

Kennan, John and James R. Walker. 2011. The effect of expected income on individual migration decisions. *Econometrica* 79(1): 211–251.

Klepeis, Neil E., William C. Nelson, Wayne R. Ott, John P. Robinson, Andy M. Tsang, Paul Switzer, Joseph V. Behar, Stephen C. Herng, and William H. Engelmann. 2001. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology* 11: 211–252.

Kline, Patrick and Enrico Moretti. 2014. People, places, and public policy: Some simple welfare economics of local economic development policies. *Annual Review of Economics* 6: 629–662.

MacDonald, John S. and Leatrice D. MacDonald. 1964. Chain migration, ethnic neighborhood formation and social networks. *Milbank Memorial Fund Quarterly* 42(1): 82–97.

Manski, Charles F. and Steven R. Lerman. 1977. The estimation of choice probabilities from choice based samples. *Econometrica* 45(8): 1977–1988.

Matz, Carlyn J., David M. Stieb, Karelyn Davis, Marika Egyed, Andreas Rose, Benedito Chou, and Orly Brion. 2014. Effects of age, season, gender and urban-rural status on time-activity: Canadian Human Activity Pattern survey 2 (CHAPS 2). *International Journal of Environmental Research and Public Health* 11(2): 2108–2124.

McFadden, Daniel L. 1973. Conditional logit analysis of qualitative choice behaviour. In Paul Zarembka (ed.) *Frontiers in Econometrics*. New York NY: Academic Press, 105–142.

McKelvey, Richard D. and William Zavoina. 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4(1): 103–120.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. Internal migration in the United States. *Journal of Economic Perspectives* 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, Joan. 2018. Economic shocks and internal migration. Processed, Universitat Pompeu Fabra.

Moretti, Enrico. 2011. Local labor markets. In Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics*, volume 4. Amsterdam: Elsevier, 1237–1313.

Munshi, Kaivan and Mark Rosenzweig. 2016. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review* 106(1): 46–98.

Patacchini, Eleonora and Tiziano Arduini. 2016. Residential choices of young Americans. *Journal of Housing Economics* 34: 69–81.

Roback, Jennifer. 1982. Wages, rents, and the quality of life. *Journal of Political Economy* 90(6): 1257–1278.

Rosen, Sherwin. 1979. Wage-based indexes of urban quality of life. In Peter N. Miezkowski and Mahlon R. Straszheim (eds.) *Current Issues in Urban Economics*. Baltimore, MD: Johns Hopkins University Press, 74–104.

Schmidheiny, Kurt. 2006. Income segregation and local progressive taxation: Empirical evidence from switzerland. *Journal of Public Economics* 90(3): 429–458.

Stopczynski, Arkadiusz, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. 2014. Measuring large-scale social networks with high resolution. *PLoS ONE* 9(4).

Topa, Giorgio and Yves Zenou. 2015. Neighborhood and network effects. In Gilles Duranton, J. Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5. Amsterdam: Elsevier, 561–624.

UNESCO and World Bank. 2018. *Culture in City Reconstruction and Recovery*. Paris: United Nations Educational, Scientific and Cultural Organization (UNESCO).

Wang, Dashun, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. 2011. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association for Computing Machinery, 1100–1108.

Zignani, Matteo, Christian Quadri, Sabrina Gaito, and Gian Paolo Rossi. 2015. Calling, texting, and moving: multidimensional interactions of mobile phone users. *Computational Social Networks* 2(13).

Appendix A. Descriptive statistics

Table A.1: Descriptive statistics at original residence

	All individuals		Movers	
	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(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 (# 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 A.2: Descriptive statistics for movers

	At new location		Mean across locations	
	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(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 A.3: 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

Appendix B. Additional tables

Table B.1: Probability of changing residential location — Robustness using number of contacts

	Dep. var.: Probability of changing residential location			
	Linear prob. model		Logit	
	(1)	(2)	(3)	(4)
Number of contacts				
0–10 min	-0.259*** (0.004)	-0.272*** (0.004)	-0.094*** (0.002)	-0.094*** (0.002)
10–20 min		-0.054*** (0.005)		-0.006*** (0.002)
20–30 min		-0.016* (0.007)		0.005* (0.002)
30–40 min		-0.023** (0.009)		-0.003 (0.003)
Degree centrality (# contacts)	0.092*** (0.001)	0.113*** (0.003)	0.023*** (0.000)	0.024*** (0.001)
Long-term resident	-0.801*** (0.022)	-0.798*** (0.022)	-0.335*** (0.009)	-0.334*** (0.009)
Language, age, gender	Yes	Yes	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Avg. probability	4.9%	4.9%	4.7%	4.8%
R^2	0.015	0.016	-	-
Pseudo R^2	-	-	0.158	0.158
N	2,136,093	2,136,093	2,136,093	2,136,093

Notes: This table replicates table 3 in the main text, replacing the share of contacts within given distances with the number of contacts within those distances.

Table B.2: Residential location choice— Robustness using number of contacts

	Dep. var.: Prob. of choosing a location conditional on moving					
	Linear prob. model			Conditional logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Number of contacts						
0–10 min	0.795*** (0.001)	0.756*** (0.001)	0.754*** (0.001)	0.808*** (0.004)	0.743*** (0.004)	0.625*** (0.003)
10–20 min	0.004*** (0.001)	0.005*** (0.001)		0.443*** (0.003)	0.414*** (0.003)	
20–30 min	-0.021*** (0.000)	-0.020*** (0.000)		0.196*** (0.004)	0.186*** (0.004)	
30–40 min	-0.021*** (0.000)	-0.020*** (0.000)		0.011** (0.004)	0.010* (0.004)	
Within same employment area	0.246*** (0.002)	0.236*** (0.002)	0.195*** (0.002)	1.016*** (0.015)	1.035*** (0.015)	1.543*** (0.013)
Within preferred language region	0.037*** (0.003)	0.042** (0.003)	0.026*** (0.003)	0.512*** (0.030)	0.569*** (0.030)	0.808*** (0.029)
Return migration		15.361*** (0.036)	15.388*** (0.036)		2.449*** (0.026)	2.714*** (0.027)
Individual x location controls	No	Yes	Yes	No	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.021	0.028	0.028	-	-	-
Pseudo R^2	-	-	-	0.220	0.243	0.206
N	25,555,189	25,555,189	25,555,189	25,342,595	25,342,595	25,342,595

Notes: This table replicates table 4 in the main text, replacing the share of contacts within given distances with the number of contacts within those distances.

Table B.3: Residential location choice — Robustness

Dep. var.: Prob. of choosing a location conditional on moving	Gap 3 Months			Gap 8 Months
	Sample 40 min	Sample 40 min + 100	Full Sample	Full Sample
	(1)	(2)	(3)	(4)
Share of contacts				
0–10 min	8.855*** (0.015)	8.919*** (0.067)	8.910*** (0.006)	8.107*** (0.014)
10–20 min	0.909*** (0.007)	0.902*** (0.013)	0.901*** (0.003)	0.949*** (0.006)
20–30 min	0.045*** (0.005)	0.043*** (0.005)	0.043*** (0.002)	0.061*** (0.005)
30–40 min	-0.038*** (0.004)	-0.024*** (0.002)	-0.024*** (0.001)	-0.009 (0.004)
Within same employment area	0.105*** (0.002)	0.067*** (0.001)	0.067*** (0.001)	0.063*** (0.002)
Within preferred language region	0.010** (0.003)	-0.002*** (0.000)	-0.001*** (0.000)	-0.002* (0.001)
Return migration	15.318*** (0.036)	13.868*** (0.337)	14.181*** (0.014)	14.074*** (0.035)
Individual x location controls	Yes	Yes	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes
R^2	0.030	0.028	0.028	0.025
N	25,555,189	30,432,919	154,175,809	24,101,887

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 4. (1) Baseline. In column (2) we restrict the sample for each mover to locations where she/he has at least one contact within 40 minutes and sample 100 random location alternatives without contact; to account for the sampling, we attach the individual specific weights to the sampled alternatives. In column (3) we use the full set of locations, in (4) we restrict the sample to locations, where she/he has at least one contact within 40 minutes but extend the gap between network formation and moving date to 8 months. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

Table B.4: Location decision and local characteristics – Robustness

Dep. var.: Prob. of choosing a location conditional on moving	Sample 40 min		Sample 40 min + 100		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Share of contacts within 0–10 min	9.038*** (0.019)	6.620*** (0.016)	9.102*** (0.086)	6.226*** (0.063)	9.128*** (0.007)	6.195*** (0.006)
Share of contacts within 0–10 min						
× housing turnover	-238.758*** (1.623)		-246.107*** (6.531)		-247.089*** (0.647)	
× childcare slots	6.724*** (0.243)		7.145*** (0.975)		7.009*** (0.096)	
× cultural events	0.042*** (0.001)		0.042*** (0.003)		0.041*** (0.000)	
× recent crimes	-1.603*** (0.014)		-1.633*** (0.061)		-1.633*** (0.006)	
× location fixed effect		13.601*** (0.032)		18.529*** (0.305)		18.908*** (0.302)
Share non-contact movers from same origin	78.734*** (0.478)	80.333*** (0.475)	73.959*** (2.228)	75.693*** (2.227)	72.472*** (0.180)	79.761*** (0.179)
Individual x location controls	Yes	Yes	Yes	Yes	Yes	Yes
Location fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.033	0.038	0.031	0.036	0.031	0.036
N	25,538,167	25,555,189	29,925,434	29,943,825	146,821,080	146,821,080

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 4. Within same employment area, within preferred language region, and return migration are included but coefficients not reported. In columns (1) and (2) we restrict the sample for each mover to locations, where she/he has at least one contact within 40 minutes (as in the benchmark specifications). In columns (3) and (4) we additionally sample 100 random location alternatives without contact within 40 minutes; to account for the sampling, we attach the individual specific weights to the sampled alternatives. In column (5) and (6) we use the full sample of all location alternatives. ***, **, and * indicate significance at the 0.1, 1, and 5 percent levels.

Appendix C. Inferring family ties

In section 8, 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 Figure 3.1, we identify six different types of potential family clusters, which we order along the following hierarchy:

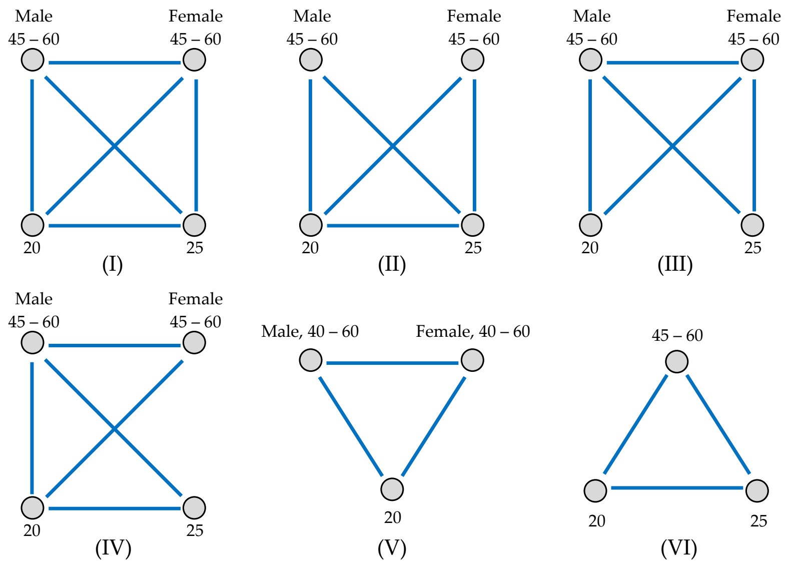


Figure C.1: Hierarchy of Family Call Patterns, Quads & Triangles.

(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 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.