Journal of Urban Economics 119 (2020) 103276

Contents lists available at ScienceDirect

Journal of Urban Economics

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

Cities and the structure of social interactions: Evidence from mobile phone data☆

Konstantin Büchel, Maximilian v. Ehrlich*

University of Bern, Department of Economics and Center for Regional Economic Development, Schanzeneckstrasse 1, 3001 Bern, Switzerland

ARTICLE INFO

JEL classification:
R10
R23
D83
D85
Z13

Keywords:
Social interactions
Mobile phones
Face-to-Face interactions
Cities
Spatial sorting

ABSTRACT

The impact of telecommunication technologies on the role of cities depends on whether these technologies and face-to-face interactions are substitutes or complements. We analyze anonymized mobile phone data to examine how distance and population density affect calling behavior. Exploiting an exogenous change in travel times as well as permanent relocations of individuals, we find that distance is highly detrimental to link formation. Mobile phone usage significantly increases with population density even when spatial sorting is accounted for. This effect is most pronounced for local interactions between individuals in the same catchment area. This indicates that face-to-face interactions and mobile phone calls are complementary to each other, so that mobile phone technology may even increase the dividends of density.

1. Introduction

Social interactions are pivotal for the diffusion of information and they directly contribute to well-being. A central feature in many models of urban economics is that a dense concentration facilitates social interactions and thereby benefits learning and productivity (for an overview see Duranton and Puga, 2004). These models build on the common assumption that distance induces costs to interpersonal exchange. The widespread adoption of information and telecommunication technologies popularized the “death-of-distance” hypothesis (e.g. Cairncross, 2001), which raises the question of whether these technologies will change the structure of cities (see Ioannides et al., 2008) or even make them obsolete. As argued by Gaspar and Glaeser (1998), the impact of mobile telecommunication technologies on the role of cities crucially depends on whether face-to-face meetings and phone calls act as *substitutes* or *complements* to each other – in the latter case, mobile phone technology may even increase the dividends of density.

This paper presents evidence that allows to assess whether face-to-face meetings and social interactions via mobile phones are complements or substitutes. We estimate the effects of distance and population density on the phone usage of individuals at different locations. Our analysis builds on anonymized call detail records *(CDRs)* which combine information about communication patterns and place of residence. The dataset covers millions of calls and text messages over a period of 12 months and allows us to address three sets of questions: *First,* how does the travel distance between two agents affect the likelihood that they interact via their mobile phones? *Second,* how does population density affect the phone usage of agents measured via call frequency, call duration, and number of unique contacts? *Third*, does density affect the pattern of *local* phone calls, that is those calls within the caller's catch-

☆ We benefited from invaluable comments by the Editor, two anonymous referees, Juan Becutti, Sascha Becker, Aymo Brunetti, Shane Greenstein, Fabian Gunzinger, Christian Hilber, Blaise Melly, Luigi Pascali, Diego Puga, Esteban Rossi-Hansberg, Elisabet Viladecans-Marsal, Alex Whalley, and Yves Zenou. We further received insightful feedback from participants at the Verein für Socialpolitik Meeting in Tellow, the Meeting of the Urban Economics Association in Minneapolis, the CRED-Workshop in Bern, the SERC Conference in London, the Applecross Conference, the SSES Annual Congress in Lausanne, the NBER Summer Institute 2017, the 6th Lindau Meeting on Economic Sciences, the 2017 SMU Conference on Urban and Regional Economics, the Meeting of the Urban Economics Association in New York and research seminars at the Paris School of Economics, the University of Milan, the IEB Barcelona, and the University of St. Gallen. A very special thanks is due to *Swisscom AG* for providing the facilities and data to conduct this research project; we are particularly indebted to Imad Aad and Holger Müller, who have accompanied the project since its start. We also want to thank *search.ch* and *comparis.ch* for providing data on travel times and usage statistics of messenger apps.

* Corresponding author.
E-mail addresses: konstantin.buechel@vwi.unibe.ch (K. Büchel), maximilian.vonehrlich@vwi.unibe.ch (M.v. Ehrlich).

https://doi.org/10.1016/j.jue.2020.103276
Received 14 January 2019; Received in revised form 26 April 2020
Available online 20 August 2020

ment area? The absence of significant distance costs and a negative impact of density on phone usage may be interpreted as supportive evidence for the "death-of -distance" hypothesis and the idea that urban centers lose ground as hubs for social exchange. Contrary, a penalty on distance, a higher phone usage and more local calls in cities would point towards a complementarity between the two modes of communication and hence contradict the hypothesis that modern telecommunication undermines agglomeration forces that operate via social interactions.

Systematic sorting complicates the analysis of our three guiding questions. Individuals with similar (unobservable) characteristics potentially cluster in space and are more likely to interact with each other. Moreover, preferences for density and phone usage may be correlated. To address these endogeneity concerns, we use individuals who permanently relocate (referred to as movers) to back out time-constant unobservables and identify density-related effects. This identification strategy relates to approaches quantifying the earning advantages of cities (e.g. Combes et al., 2008). As movers may share characteristics that differ systematically from the average, we discuss heterogeneity of the effects and compare the results for movers to those identified for stayers exploiting the plausibly exogenous revision of public transport schedules. This revision changed the effective distance between individuals and allows us to compare the calling behavior of individuals before and after the public transport timetable was updated.

Several studies examine the use of communication technology from a spatial perspective. They document an increase in phone usage with city size (see Gaspar and Glaeser, 1998; Charlot and Duranton, 2006; Schläpfer et al., 2014) or illustrate a strong negative correlation between distance and the likelihood for an exchange via email and Facebook (e.g. Bailey et al., 2018, 2020; Levy and Goldenberg, 2014). However, neither of these contributions address non-random sorting.[1] We attempt to separate the causal effects of distance and density from sorting by tracing changes in phone usage over time. This strategy allows us to perform a comprehensive set of tests to discriminate between complementarity and substitutability of phone and face-to-face interactions. Density externalities tend to be highly localized (e.g. Arzaghi and Henderson, 2008; Shoag and Veuger, 2018). CDRs are especially suitable to study such effects due to their fine-grained spatial resolution. They have been used in recent research, for instance, to analyze correlations between calling patterns and regional economic development (Eagle et al., 2010) to study solidarity in the context of natural disasters (Blumenstock et al., 2016) or to examine the role of referrals in labor markets (Barwick et al., 2019).

We demonstrate that social interactions via phones are markedly localized, with the distance-elasticity of link formation falling quickly and converging to zero at about 100 minutes travel distance. How do these distance related costs affect phone usage at locations of different population density? We show that the overall intensity of phone interactions measured by call frequency and duration increases with population density, while the number of unique contacts remains unaffected when we account for sorting. The impact of density is even more pronounced for local interactions, as individuals in cities display a higher number of unique contacts within their neighborhood, and call those nearby more frequently and for a longer duration than individuals living in sparsely populated areas. The elasticities of local interactions in terms of frequency, duration, and unique number of contacts with regard to population density ranges between 0.1 and 0.4. These findings remain robust when we account for unobservable place characteristics by instrumenting population density, when we include text messages or when we examine different subsamples of the population. Our results support the hypothesis that mobile phone calls complement (not substitute) face-to-face interactions.

2. Testing the complementarity between mobile phone and face-to-face interactions

In this section, we briefly motivate three sets of tests that allow us to assess whether phone and face-to-face interactions are complements or substitutes.

The *first* test examines the impact of geographical distance on the likelihood that two individuals interact via their phones: Assume that face-to-face interactions are subject to distance costs between individuals. If face-to-face interactions and telecommunication are substitutes, then people who live in close proximity would need to call each other less often, since they can easily meet face-to-face. In contrast, a large distance penalty would point towards a complementarity between the two modes of interaction, as individuals who live in the same neighborhood are also more likely to call each other.

For the *second* set of tests, think of a setting similar to Gaspar and Glaeser (1998). An individual first decides whether to carry out an activity privately or jointly with a peer. Conditional on interacting, the individual then decides whether to carry out the joint activity in a face-to-face meeting or via an electronic interaction such as a phone call. While the value of face-to-face interactions is strictly higher than the value of phone calls, face-to-face meetings are associated with distance costs depending on the physical location of the peer. The optimal choice between the two modes of interaction depend on the activity-specific quality of the match between the individuals which has an idiosyncratic component.[2] Cities, due to their high population density, have lower distance costs, which has two effects: First, it raises the expected net benefits of interacting such that individuals find it in more circumstances optimal to carry out an activity jointly. Second, with lower distance costs, the mix of interactions shifts from electronic towards face-to-face interactions. If cities were to display a higher absolute level of phone interactions than low density places, this would imply that the first effect dominates and that the total number of interactions – via phone and face-to-face – is higher in cities. We thus estimate effects of population density on individual-level measures of phone usage, i.e. the frequency and duration of phone calls between individuals as well as the number of unique contacts within an individual's network.

The *third* set of tests builds on the same measures but recalculates them for calls within the catchment area of each individual's residence, which we define to cover a perimeter of 15 minutes car travel time. City residents have an innately larger number of potential contacts within their catchment area. Thus, complementarity between face-to-face interactions and phone calls is especially bound to be reflected in a more localized network, that is a larger number of unique phone contacts, a higher frequency of calls, and a longer call duration within one's neighborhood.

Before we lay out the empirical strategy and results for the first set of tests in Section 4, and for the second and third sets of tests in Section 5, we next describe the data set.

3. Data

The main dataset used in this paper is provided by Switzerland's largest telecommunications operator, Swisscom AG, whose market share for mobile phones is 55 percent (ComCom, 2015). The market share spreads relatively homogeneous across space, with an interquartile range of 47 percent to 62 percent across municipalities. The data comprises CDRs of all calls made by the operator's customers between June 2015 and May 2016. The CDRs include the anonymized phone

[1] Other studies examine the importance of geographic proximity for the formation of friendship among students (e.g. Marmaros and Sacerdote, 2006; Kim et al., 2017). While these studies address endogeneity, they allow no conclusions regarding the role of telecommunication technologies.

[2] The idiosyncratic component reflects, for instance, activities that constrain their time or appetite for interactions.

Table 1
Phone Usage and Sociodemographics of Private Mobile Phone Customers.

	Mean	SD	Min	Max
Monthly Phone Usage, June 2015 – May 2016 (pooled)				
Number of Calls	32.332	37.286	1	1589
Number of *Local* Calls	21.641	32.250	0	1551
Duration of Calls (Minutes)	117.176	167.929	0.17	3292
Duration of *Local* Calls	62.187	127.214	0	3065
Number of Unique Contacts	9.214	7.874	1	470
Number of Unique *Local* Contacts	7.074	7.221	0	208
Sociodemographics				
Age	34.964	13.561	20	60
Female	0.522	–	0	1
Language: German	0.681	–	0	1
Language: French	0.270	–	0	1
Language: Italian	0.043	–	0	1
Language: English	0.006	–	0	1

Notes: Local refers to the subset of calls within a radius of 15 minutes around an agent's residence. The table is based on the subsample of customers with phone activity in all 12 months, which we also use in the main analysis (*N*=866,646). Further filters as described in Section 3.

number of caller and callee, a date and time stamp, a binary indicator for private and business customers, a code for the type of interaction (e.g. call, SMS, MMS), the duration of calls in seconds, and the x-y-coordinates of the caller's main transmitting antenna. We observe finely grained information on about 15 million calls and text messages per day, covering about 9.1 million phones, of which 4.1 million are mobile phones and 2.7 million are private mobile phones.

Phone data seems particularly appropriate to study social interactions, since most people use some combination of calls, direct encounters, and text messages to communicate. Moreover, mobile phone plans in Switzerland typically include unlimited domestic calls and differentiate primarily based on the amount of data included. Thus, voice calls involve a zero marginal cost and, as recent survey data shows, they have "remained popular in Switzerland despite the onslaught of messaging services."[3]

Along with the anonymized CDRs, the operator also provided monthly updated *customer information* including billing address, language of correspondence (German, French, Italian, English), age and gender. Table 1 summarizes the socio-demographic characteristics of mobile phone customers in our sample, while Table A.3 shows correlations between census and customer data for various subpopulations. The comparison suggests that the data is highly representative of the Swiss population even at very local levels.

The anonymity of customers was guaranteed at all steps of the analysis. We never dealt with or had access to uncensored data. A data security specialist retrieved the CDRs from the operator's database and anonymized the telephone numbers using a 64-bit hash algorithm that preserved the international and local area codes. He further removed columns with information on the transmitting antenna before making the data available. Once the anonymized data were copied to a fully sealed and encrypted workstation, we ran the analysis on site. To utilize information on the transmitting antenna we passed scripts to the operator's personnel who executed them for us.

3.1. Observing social interactions in phone data

Our primary aim is to observe social interactions, but not every instance of phone activity qualifies as such in the narrower sense so that the dataset needs to be cleaned beforehand (for a discussion see Blondel et al., 2015). In our benchmark analysis, we filter the data as follows: *First*, we restrict the analysis to calls between mobile phones. Mobile phones are personal objects and are thus representative of the social network of a single person, while calls from landlines possibly resemble overlapping social networks as they are usually shared by multiple users. We assume that landline customers also own a mobile phone; hence we do not miss relevant links, while including calls to both devices may artificially inflate the size of a person's network. For the same reason, all results are based on customers who have registered only one active mobile phone number. Customers with multiple active numbers typically include corporate customers, as well as parents acting as invoice recipients for their children. *Second*, we limit the analysis to outgoing calls in order to cover intra-operator and inter-operator activity equally well and to filter out promotional calls by call centers. *Third*, calls with a duration of less than 10 seconds are considered accidental and are therefore excluded from the analysis. *Fourth*, we drop mobile phone numbers that display implausibly low or high monthly usage statistics, with a minimum threshold of 1 minute and a maximum threshold of 56 hours per month. This removes inactive numbers as well as commercially used phones. *Fifth*, the analysis is limited to private mobile phones, so that business calls between corporate customers do not create noise in our measures. *Sixth*, some specifications require address information for both caller and callee such that inter-operator calls cannot be used in all steps of the analysis. Those estimations are therefore based on intra-operator calls, which we weight according to the operator's market share at the callee's billing address. *Finally*, we only use the first 28 days of each month to make the data comparable across different months.

These steps eliminate approximately 37 percent of the total duration of phone calls recorded, leaving us with around 60 million calls per month that amount to a total duration of 200 million minutes (for details see Table A.1 in the appendix). We assess the robustness of our insights by adjusting our filtering procedure. In particular, we re-examine the results of our benchmark estimations for phone usage measures incorporating text messages and phone usage measures based on both incoming and outgoing calls.

3.2. Descriptive statistics on phone usage and local characteristics

Table 1 shows summary statistics on mobile phone usage for customers aged 15 to 64.[4] These include measures that inform about the number of unique contacts, about the intensity of interactions measured by frequency and duration of calls, and whether the call occurs within the local catchment area of the caller. The local catchment area is defined as those postcodes that are located within a 15 minutes road travel time perimeter from the individual's place of residence. The average private mobile phone user makes 1.2 calls per day with a cumulative duration of four minutes to a bit more than nine unique contacts. Roughly 65 percent of the phone usage matches our definition of local calls, i.e. calls to recipients living within a 15 minutes car travel perimeter from the caller's address.

Table A.2 in the appendix displays the corresponding statistics by age, gender, and type of residence. A few observations stand out: Female and older customers display on average a lower calling frequency and number of unique contacts than male and younger customers. Although they make on average fewer calls to a smaller number of unique contacts, the cumulative duration of calls is higher for female than for male customers. Finally, the youngest customers (i.e. 15–24 year) maintain a more local call network compared to older customers. Fig. A.1 in the appendix further shows that the distributions of phone usage measures are markedly right-skewed and that the network uncovered by the data exhibits characteristic features of other socially generated networks documented in the literature (see Jackson and Rogers, 2007; Watts, 1999):

[3] See *This Is How Swiss Make Phone Calls* by Moneyland.ch, https://www.moneyland.ch/en/switzerland-telephone-call-survey-2018 (last access: 09.01.2019).

[4] Due to privacy concerns, we worked with decimal age-brackets. This means that a customer aged 24 was assigned to the 20-bracket, while a customer aged 25 belongs to the 30-bracket.

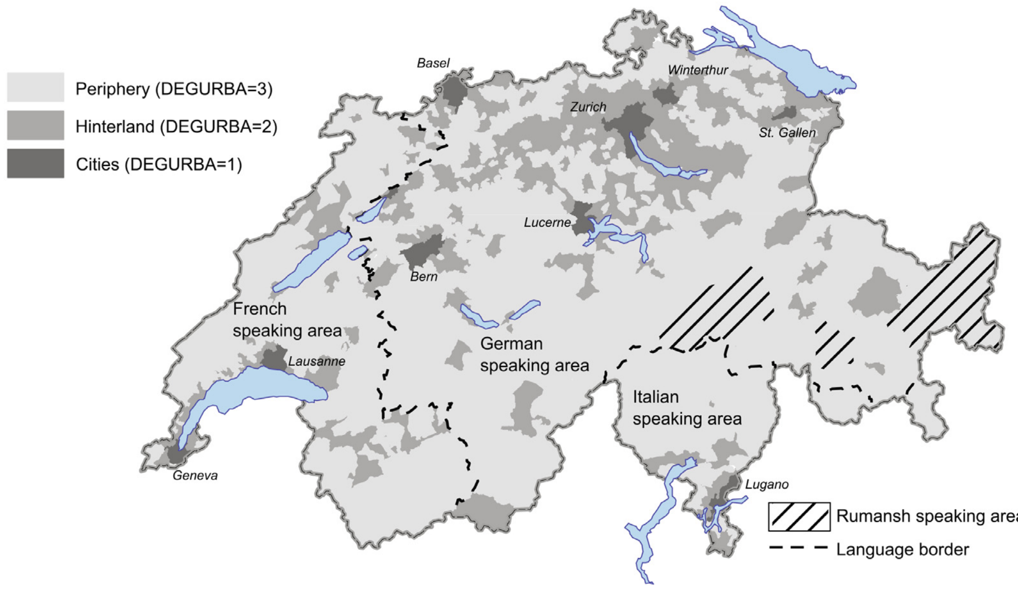


Fig. 1. Degree of Urbanization – Cities, Hinterland and Periphery.

A short average path length between pairs and "fat tails" in the distribution of the number of unique contacts.

The phone data are complemented by various municipal statistics provided by the Federal Statistical Office (FSO), including population figures and the degree of urbanization as classified by EUROSTAT.[5] Fig. 1 shows the regional variation in urbanization based on the aforementioned measure. We map the municipal statistics on the areal boundaries of postcodes. If the area of a postcode intersects with several municipalities, we weight the municipal statistics based on the stock of buildings as of 2015; for categorical variables the postcode is assigned to the category with the highest accumulated weight. We also compute geographical distances between pairs of postcodes using GIS software. Car driving distances between centroids of postcodes and public transport travel times for all existing pairs of stops were obtained from *search.ch*. Table 2 displays the descriptive statistics for our cleaned sample of 3152 postcodes.

4. Distance and social interactions via mobile phone

4.1. Empirical model

We consider a directed network with N nodes each representing a unique phone customer which we denote by $i \in \mathcal{N} = \{1, \dots, N\}$. A link between individuals i and j is defined by $g_{ij} = 1$, while the absence of a link is marked as $g_{ij} = 0$. This network can then be characterized by a pair $(\mathcal{N}, \mathcal{G})$ where $\mathcal{G} = [g_{ij}]$ is a $N \times N$ adjacency matrix. We observe the network's adjacency matrix $\mathcal{G}_t = [g_{ij,t}]$ in each month $t \in \{1, \dots, 12\}$. We assume that rational agents i and j establish a link if the net surplus from doing so is positive (c.f. Graham, 2015). Accordingly, we specify the linear probability model of two individuals i and j forming a link as

$$g_{ij,t} = T'_{ij,t}\eta_1 + F'_{ij,t-1}\eta_2 + Z'_{ij}\rho + \phi_1 D_{i,t-1} + \phi_2 D_{j,t-1} + m_{ij} + U_{ij,t}, \tag{1}$$

where vector $T_{ij,t}$ measures the distance between i and j, $F_{ij,t-1}$ reflects the number of contacts i and j have in common, Z_{ij} is a vector of dyad-specific but time-invariant covariates, $D_{i,t-1}$ and $D_{j,t-1}$ capture differences in sociability based on both parties' number of unique contacts, m_{ij} captures unobserved pair specific heterogeneity, and $U_{ij,t}$ denotes the random utility component which is assumed to be i.i.d. with mean zero.[6]

The distance measures represented by vector $T_{ij,t}$ comprise the log travel time between individual i's and j's residence as well as a dummy for same workplace.[7] We use road travel times as well as public transport travel times. The place of work dummy equals one if they predominantly use transmitting antennas within the same 5 km radius during business hours. We discretize the number of common friends, such that we obtain two dummy variables contained in $F_{ij,t-1}$: The first indicator equals one, if agents i and j share at least one common social contact, while the second indicator equals one if agents i and j share at least two common contacts.[8] The dyad-specific covariates in vector Z_{ij} include three dummy variables indicating same age (i.e. same decimal age bracket), same gender and same language.

The dyad-specific fixed effect in Eq. (1) accounts for matching based on unobservables which may bias estimates of pooled OLS models without fixed effect. If individuals with common unobservable attributes are more likely to cluster regionally and thus live closer together, our distance measure will be negatively correlated with the error term. Including the dyad-specific fixed effect will take out time invariant factors that affect the matching quality. Our causal identification builds on two sources of changes in effective distance between agents, i.e. changes in travel time T_{ij} which should lead to adjustments in the calling behavior.

[5] See http://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_DEGURBA (last access: 01.06.2016) for more information on the EUROSTAT DEGURBA measure.

[6] Note that the number of mutual contacts, $F_{ij,t-1}$, enters with a lag. This implies that agents form/dissolve links myopically, as if all features of the previous period's network remain fixed. Assuming this structure eliminates contemporaneous feedback, which would confound inference (Graham, 2015).

[7] We have estimated the models also with geographical distance instead of travel time which does not qualitatively affect our results. However, due to the rugged landscape in Switzerland travel time is the more relevant measure. Bailey et al., 2020 show that even in a dense urban environment travel time and travel costs are substantially stronger predictors of Facebook ties than geographic distance.

[8] We discretize the number of mutual friends, because the continuous measure yields imprecise (yet significant) estimates. Sensitivity checks showed diminishing effects of mutual friends as mutual friends beyond two did not significantly add to the link likelihood.

Table 2
Main Descriptive Statistics across Postcodes.

	Mean	SD	Min	Max
Area in km^2	12.927	19.215	0.014	242.904
# Customers within catchment area (i.e. wihtin 15 min. travel time perimeter)	14,683	16818.31	50	107,549
Distance: Postcode i to j				
Euclidean Distance (km)	111.931	59.501	0.336	353.852
Travel Time by Car (min.)	142.804	69.033	0.283	453.508
Travel Time by Public Trans. (min.)	269.904	100.181	1.008	713.000
Degree of Urbanization				
City	0.035	–	0	1
Periphery	0.336	–	0	1
Hinterland	0.629	–	0	1
Main Language				
German	0.628	–	0	1
French	0.295	–	0	1
Italian	0.065	–	0	1
Rhaeto-Romanic	0.012	–	0	1

Sources: The sample covers 3152 postcodes. Areal data from Swisstopo; population data, language shares, and degree of urbanisation from the Federal Statistical Office; car travel times from *search.ch*; number of customers from Swisscom. Data from postcodes with less than 50 customers were deleted due to data privacy requirements.

The first stems from relocation of individuals and the second is a natural experiment caused by a major public transport investment which allows us to focus on the link formation of stayers.

A practical issue that arises with estimating the outlined link formation models is the size of the adjacency matrix that potentially includes $(2 \cdot 10^6)^2$ unique pairs of individuals. It is neither computationally feasible to estimate the models based on all these pairs nor necessary for obtaining consistent estimates of the parameters of interest as is shown by Manski and Lerman (1977), and Cosslett (1981). Since we know the true number of potential and established links we can use a stratified sample and adjust the estimates with the respective sampling weights. Our choice-based sample results from an endogenous stratified sampling scheme where each stratum is defined according to the individual responses, that is the binary values taken by the response variable $g_{ij,t}$.[9] This sampling structure requires the availability of prior information on the marginal response probabilities which is in our setting available due to the observation of $\mathcal{G}_t$.

4.2. Main results

We first examine the relation between distance and mobile phone calls by plotting the share of links against the share of potential links by radius. Fig. 2 illustrates the rapid decline of mobile phone interactions across space: Almost 50 percent of links are formed within a 5 km perimeter that covers on average less than 1 percent of the population.

4.2.1. Estimates based on distance changes due to moving

Fig. 2 neither accounts for biases due to spatial sorting of similar types nor is it informative about the relative importance of distance. We therefore proceed to the link formation models, outlined in the previous section and analyze the link formation of movers which allows us to back out unobserved pair-specific factors. Before we elaborate on the results for the distance changes due to moving, let us briefly discuss whether the movers in our dataset are representative of the movers in the Swiss population. While we observe somewhat more movers in the raw phone data, i.e. about 6 percent, compared to 4.2 percent in the postal data, Table 3 shows that the moving patterns match very well. Another concern may be that movers are systematically different from non-movers. Table A.4 in the appendix compares phone usage statistics and socio-demographics between movers and stayers. While movers are considerably younger than non-movers, they only marginally differ along the other dimensions such as calling frequency, duration and number of unique contacts. In terms of gender and language, movers are practically identical to non-movers.

Table 3
Representativeness of Movers Recorded in Mobile Phone Data.

	Sample	Postal Data
Percent of movers by distance		
0–30 min	71.75	70.62
> 30 min	28.25	29.38
Percent of movers by DEGURBA classification		
city to hinterland/periphery.	9.90	9.89
hinterland/periphery to city	13.75	12.60
within hinterland/periphery	20.79	18.15
no change	55.56	59.36

Notes: Column (1) reports moves based on changes in the postcode of the billing address for cellphone users in our filtered sample for June 2015-May 2016. Column (2) reports moves based on address changes recorded by Post AG, the Federal postal service, for January-December 2014.

In Table 4 we estimate the effect of distance on link formation. Columns (1) and (2) report pooled OLS models based on a 5 percent random sample of private mobile phone customers;[10] column (1) only includes car travel time, while column (2) estimates the benchmark model as in Eq. (1) without controlling for unobserved dyad-specific heterogeneity. Columns (3) to (6) show the results of pooled OLS models that were estimated based on all private mobile phone customers that moved between June 2015 and May 2016; columns (3) to (5) use car travel time to account for the distance between the homes of two individuals, while

[9] The main motivation behind this approach is usually the possibility of oversampling rare alternatives, which can improve the accuracy of the econometric analysis but also reduce survey costs. However, in our case we undersample those dyads with $g_{ij,t} = 0$ in order to enhance computational efficiency.

[10] Estimating the model for all customers and their potential interactions is computationally not feasible. The estimates are therefore based on a random sample and proved very robust to sampling.

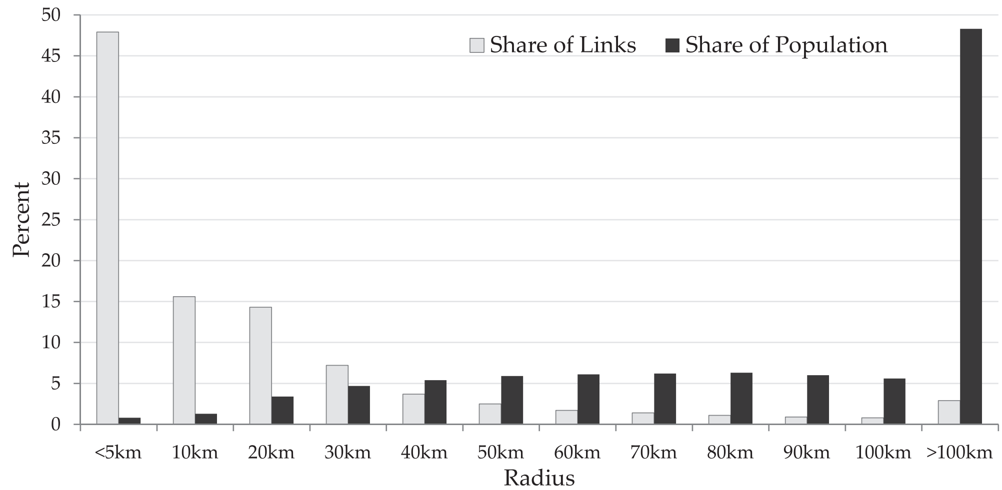


Fig. 2. Share of Links Measured via Mobile Phone Calls and Population by Radius *Notes:* The share of links reflect mobile phone calls made in June 2015. The radius is calculated based on the distance between the caller's and callee's place of residence. Population statistics comprise number of mobile phone customer's by postcode.

column (6) is based on public transport travel times. Columns (7) to (10) are estimated with the same sample of movers as columns (3) to (6) but include pair fixed effects to absorb time constant dyad-specific heterogeneity. Coefficients were multiplied by 10,000 an can be interpreted as basis points. Hence, a coefficient equaling one translates to a marginal increase in $Pr(g_{ij}=1)$ of a hundredth percentage point.

In all estimated specifications the travel time between two agents enters negatively, implying that distance is indeed costly when forming and maintaining a link. To compare the estimates across different samples we compute distance elasticities adjusted for base probabilities i.e. the relative marginal effect $\frac{\hat{\eta}_1}{Pr(g_{ij}=1|T_{ij}=0)}$. Comparing the distance elasticities adjusted for base probabilities in the most parsimonious fixed effect specification (column 7: $-0.020/0.116=-0.172$) to the equivalent pooled OLS specification (column 3: $-0.147/0.718=-0.204$) suggests that unobservable matching attributes are correlated with travel distances between i and j, so that the plain OLS estimate overstate the distance penalty to interactions via phone.

The quadratic specifications for road travel times in columns (2), (5), and (9) reveal a convex relationship between link formation and distance. According to the fixed effect estimates in column (9), doubling the distance from 10 minutes road travel to 20 minutes road travel reduces the likelihood for an interaction via phone by 48 percent.[11] The marginal effect of distance reaches zero at 92 minutes (i.e. $\exp(\frac{1.367}{2\times0.151})$ in the pooled OLS specification for movers (see column 5), and 106 minutes (i.e. $\exp(\frac{0.140}{2\times0.015}))$ when pair fixed effects are included (see column 9).

Public transport travel times, as used in columns (6) and (10), yield very similar results. According to the fixed effect estimates in column (10), doubling the distance from 10 minutes public transport travel time to 20 minutes public transport travel time reduces the likelihood for an interaction via phone by a factor of 32 percent. The marginal effect of distance reaches zero at 156 minutes (i.e. $\exp(\frac{0.101}{2\times0.010})$ when pair fixed effects are included, hence the negative effect of distance is very pronounced at small distances and relatively quickly fades out, no matter whether public transport or road travel times are used.

In order to allow for a more flexible functional form, one can replace the linear/quadratic distance functions by a series of dummies capturing bins of 5 minutes car travel time. We also estimate non-linear models to accommodate for the binary dependent variable. Section B.1 in the appendix confirms that results are robust to using a Logit model as well as modeling distance with a series of dummies.

Supporting the hypothesis of complementarity between face-to-face and phone interactions, the distance gradient is generally steeper at high density places where a large number of potential contacts live nearby. As columns (4) and (8) show, doubling the population density increases the detrimental impact of distance on the likelihood of calling each other by about 3 percent.

In addition to being neighbors, working in the same area also increases the likelihood of link formation. The coefficient for the dummy variable "Same Workplace" ranges between 0.1 and 0.5., which is roughly ten times the estimated effect of speaking the same principal language. Thus, distance in terms of both residence and workplace reduce the likelihood of interacting via phone. The coefficients for both "Common Contact" variables are highly significant. Column (4) shows that the probability of forming a link with another person increases by up to 19 percentage points, if one shares at least two common contacts. As one would expect, the estimates are considerably smaller in column (8), which controls for matching quality by employing dyad-specific fixed effects. Nonetheless, the additional link-surplus of 1.2 percentage points due to triadic relations – as obtained in the most conservative specification – is quantitatively substantial.

The pooled OLS specifications (2), (4), (5) and (6) also control for socio-demographic (dis)similarities, namely dummies for same language, same gender, and same age (decimal brackets), as well as the absolute age difference between customers i and j. Our results unambiguously point toward homophily, which is the well documented tendency of individuals to bond with similar others (e.g. McPherson et al., 2001; Currarini et al., 2009). For instance, individuals who share the same principal language are more likely to interact than individuals with different languages. The same holds true for age and gender.

The identification strategy used to obtain these results may raise the concern that movers differ systematically from the rest of the population. Apart from age, Table A.4 in the appendix shows that differences in both individual characteristics and phone usage behavior are relatively small between movers and non-movers. In addition, re-estimating the main specifications for different sub-groups of the population does not point toward relevant heterogeneity in the distance penalty (see Table B.4).

Another potential shortcoming of the above tests may be that movers differ from the population average along several *unobservable* characteristics which could also be related to phone usage. Identification of the

[11] Based column (9) in Table 4 we calculate the reduction as follows: $[(0.329-0.140\times\ln(20)+0.015\times\ln(20)^2)-(0.329-0.140\times\ln(10)+0.015\times\ln(10)^2)]/[0.329-0.140\times\ln(10)+0.015\times\ln(10)^2]$.

Table 4
Gravity Model of Social Interactions via Mobile Phones, LPM-Models.

Transport mode	All Customers, Pooled OLS		Movers, Pooled OLS				Movers, OLS with Pair Fixed Effects			
	Car (1)	Car (2)	Car (3)	Car (4)	Car (5)	Public (6)	Car (7)	Car (8)	Car (9)	Public (10)
Travel distance between individuals										
Ln(Travel Time$_{ij,t}$)	-0.165***	-1.887***	-0.147***	-0.101***	-1.367***	-0.941***	-0.020***	-0.029***	-0.140***	-0.101***
	(0.000)	(0.037)	(0.000)	(0.001)	(0.020)	(0.023)	(0.000)	(0.001)	(0.005)	(0.005)
Ln(Travel Time$_{ij,t}$)^2		0.209***			0.151***	0.090***			0.015***	0.010***
		(0.004)			(0.002)	(0.002)			(0.001)	(0.000)
Ln(Travel Time$_{ij,t}$) × Ln(PopDensity$_{i,t}$)				-0.002***				-0.001***		
				(0.000)				(0.000)		
Main control variables										
Same Workplace$_{ij,t}$		0.124***		0.494***	0.218***	0.389***		0.101***	0.100***	0.095***
		(0.024)		(0.018)	(0.018)	(0.019)		(0.002)	(0.002)	(0.001)
Same Language$_{ij,t}$		0.029***		0.044***	0.024***	0.004***				
		(0.001)		(0.001)	(0.001)	(0.001)				
Same Gender$_{ij,t}$		0.007***		0.007***	0.007***	0.006***				
		(0.001)		(0.001)	(0.001)	(0.001)				
Same Age$_{ij,t}$		0.025***		0.035***	0.035***	0.034***				
		(0.002)		(0.001)	(0.001)	(0.001)				
> 0 Common Contacts$_{ij,t-1}$		129.066***		169.179***	169.000***	163.831***		9.592***	9.592***	9.119***
		(2.812)		(3.290)	(3.290)	(3.255)		(0.281)	(0.281)	(0.275)
> 1 Common Contacts$_{ij,t-1}$		1515.627***		1886.335***	1886.203***	1845.203***		126.814***	126.814***	121.139***
		(116.663)		(123.998)	(123.987)	(125.161)		(9.037)	(9.036)	(8.910)
Const.	0.800***	4.204***	0.718***	0.516***	3.063***	2.465***	0.116***	0.155***	0.329***	0.293***
	(0.001)	(0.080)	(0.001)	(0.008)	(0.045)	(0.058)	(0.002)	(0.004)	(0.010)	(0.012)
R^2	0.001	0.037	0.001	0.045	0.045	0.044	0.533	0.533	0.533	0.536
Further Controls	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Pair FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49,172,284	49,172,284	67,964,704	67,964,704	67,964,704	65,032,940	67,964,704	67,964,704	67,964,704	65,032,940

Notes: We use monthly data for June 2015–May 2016. The *sample in columns (1) and (2)* was drawn with a 5% probability from all private mobile phones that were used every month at least once. The *sample in columns (3) to (10)* covers all movers who used their phone every month at least once. All *coefficients* are multiplied by 10000, and therefore can be interpreted as basis points. *Same Age* is unity for individuals in the same decimal age-bracket. *Further controls* include the number of unique contacts (degree centrality) for both agents and the absolute age difference between agents *i* and *j*. Standard errors in parentheses. $+p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

causal impact of distance on link formation may be impeded by *changes* in these characteristics, which would not be captured by the fixed effects. Ideally, the identification strategy based on movers would exploit an exogenous shock that displaced individuals to new locations similar as in Catalini (2017). Since such a shock did not occur in our setting, we can only analyze movers that *decided* to relocate.

In order to avoid these potentially confounding effects in the mover analysis, we re-estimate Eq. (1), but instead of using movers, we identify η_1 based on exogenous changes in the public transport timetable.[12]

4.2.2. Estimates based on distance changes following a major revision of public transport schedules

Public transportation is a frequently used mode of transportation in Switzerland, which has one of the densest railway networks worldwide.[13] After the completion of an underground cross-city route in Zurich and several new railway connections, the Swiss Federal Railways company (SBB) issued a revised timetable on 13 December 2015. This was the most substantial change of the SBB's timetable since 2004, affecting both the frequency of connections and journey times across Switzerland.[14] The largest changes occurred in the canton of Zurich and surrounding regions but all other parts of the country also displayed changes in public transport times (see Section A.5 in the appendix for more details). The planning of Switzerland's public transport schedules is highly centralized; the SBB holds a market share of 80% in rail traffic so that local providers coordinate their services with the SBB. This centralization brings about nationwide changes in public transport connections triggered by newly established connections of the federal railway. Moreover, it facilitates reliable timetable queries from webservices such as *search.ch* (our data source).

To calculate changes in the public transport travel time between two places, we use information on the quickest connections between all pairs of public transport stops including the frequency of available connections for a two hour window between 6 am and 8 am. Our measure of public transport travel times captures journey and waiting time, and is defined as

$$Public\ Transport\ Travel\ Time = \frac{120\ min.}{\#Available\ Connections} + Journey\ Time. \tag{2}$$

To obtain a comparable travel time matrix for 2017, we use the same selection of stops as in the 2015 matrix. Any changes between travel times in 2015 and 2017 can then be attributed to the timetable revision on 13 December 2015. The distribution of changes in travel times between

[12] Catalini et al., 2019 follow a related identification strategy and study whether the introduction of new routes by a low-cost airline impacted the collaboration decisions scientists. They find strong evidence for spatial frictions in scientific collaborations.

[13] For instance, public transportation covers about 60 percent of the commutes in the Zurich area. The central station of Zurich counts more than 150 million passengers per year and belongs to the top five most busiest train stations in Europe (see SBB Passagierfrequenz, 2016 at https://data.sbb.ch/explore/dataset/passagierfrequenz/table/?sort=bahnhof_haltestelle&refine.bezugsjahr=2016, last access: 09.01.2019). For a comparison of railway densities see for instance https://w3.unece.org/PXWeb/en (last access: 15.08.2019).

[14] In total, the capacity of regional rail services was expanded by 3.3 percent. See https://company.sbb.ch/de/medien/medienstelle/medienmitteilungen/detail.html/2015/11/1111-1.

Table 5
Changes in Public Transport and Social Interactions via Phone, LPM-Models.

Transport mode	Switzerland				Canton Zurich		
	Car (1)	Public (2)	Public (3)	Public (4)	Public (5)	Public (6)	Public (7)
Ln(Travel Time$_{ij,t}$)	-0.364***	-0.405***	-0.008**	-0.055	-0.380***	-0.022*	-0.326+
	(0.005)	(0.006)	(0.003)	(0.087)	(0.016)	(0.010)	(0.189)
Ln(Travel Time$_{ij,t}$)^2				0.004			0.035+
				(0.008)			(0.021)
Constant	3.046***	2.157***	0.099***	0.222	1.743***	0.178***	0.832*
	(0.042)	(0.034)	(0.017)	(0.023)	(0.073)	(0.043)	(0.423)
R^2	0.001	0.001	0.523	0.523	0.001	0.554	0.554
Pair FE	No	No	Yes	Yes	No	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Postcode Pairs	5,952,616	5,384,294	5,384,294	5,384,294	203,227	203,227	203,227
Observations	87,005,282	83,183,964	83,183,964	83,183,964	18,149,188	18,149,188	18,149,188

Notes: We use data from three-months windows prior and after the change in the public transport timetable on December 13th 2015, i.e. June 2015–August 2015 and March 2016–May 2016. The *sample* covers only non-movers (both caller and callee) who used their phone every month at least once. In column (1)–(3), we drop observations in the canton of Ticino as these were affected by an infrastructure change not recorded in our travel time data. All *coefficients* of the linear probability models are multiplied by 10000, and can be interpreted as basis points. Standard errors are clustered by postcode pair and reported in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

postcodes includes positive as well as negative values with the 10th-percentile change being a 34 minutes decrease in travel time, the median change equaling 0 minutes, and the 90th-percentile change amounting to a 12 minutes increase.

To reduce noise, we employ data from three-months windows prior and after the change in the public transport timetable, namely June–August 2015 and March–May 2016. We also restrict the sample to individuals who keep the same billing address, so that the estimates of η_1 are not affected by the potentially endogenous moving decision.

Table 5 shows the results for Swiss customers without change in the billing address (columns 1–4) and a subsample of non-moving residents of the canton Zurich (columns 5–7), where the largest changes occurred. As in the previous models, distance is negatively correlated with the probability that two agents form and maintain a link. In column (1) and (2) we compare the effect of distance as measured by road and public transport travel time for the linear, pooled OLS models: The marginal effects of distance relative to the base probabilities turn out to be similar for both transport modes with -0.12 and -0.19 for car and public transport, respectively. Once we include dyad-specific fixed effects in column (3) and identify η_1 from the exogenous change in public transport schedules, the distance elasticity relative to the base probability drops significantly. In this case the value of the elasticity relative to the base probability (i.e. -0.08) is less than half of the elasticity obtained from the model without fixed effects (i.e. -0.19). The same holds true when we compare the pooled OLS and fixed effects specifications for stayers in the Canton of Zurich, see columns (5) and (6) of Table 5. This again documents the importance of unobservable matching attributes which are related to distances and bias plain OLS estimates.

When a squared term of distance is added, the estimates' precision drop but again enter positively pointing to a convexly decreasing relation. The point estimates of Tables 4 and 5 can only be roughly compared, since Table 5 pools three consecutive months into one cross-section, which increases the base probability that a link is observed by a factor of approximately three.[15] Yet, the distance elasticities adjusted for base probabilities are quite similar: The values are -0.12 in column (5) of Table 5 and -0.17 in column (7) of Table 4. Comparing the quadratic specifications confirms that the two identification approaches yield similar distance decay functions: Column (7) of Table 5 shows that doubling the distance from 10 minutes public transport travel time to 20 minutes public transport travel time reduces the likelihood for an interaction via phone by 37 percent, and the distance elasticity converges to zero at 108 minutes. Computing these figures for column (10) in Table 4, which is based on movers, we obtain 32 percent and 156 minutes.

The public transport experiment yields consistent estimates of the distance elasticity if the induced changes in travel times are not systematically related to changes in pair-specific unobservables at the point in time when the new public transport schedule became relevant. We verify this assumption by evaluating the presence of differences in the regional trends of link formation prior to the change in the public transport timetable. First, we compare (pre-shock) sociodemographics and phone usage statistics between those individuals that reside in postcodes with below median changes in public transport travel times to individuals that live in postcodes with above median changes in public transport travel times (see Table A.5). It appears that French-speaking customers in rural areas are slightly overrepresented among those experiencing above median changes in public transport travel times, but overall the two groups share almost identical characteristics. Second, we estimate the same model specifications as in columns (6) and (7) of Table 5 assuming that the change in public transport travel times took place six months earlier. This placebo check yields insignificant estimates (see appendix B.2), and hence rejects the concern that our results are driven by regional trends in social interactions prior to the change in public transport accessibility.

We finally compare the functional form obtained from the different models in more detail. Fig. 3 plots the predicted probability for $g_{ij} = 1$ relative to the base probability at a distance of 15 minutes travel time for five models: *(i)* The pooled OLS regression based on a random sample of private mobile phone customers and car travel times (i.e. column 2 in Table 4), *(ii)* the pooled OLS regression based on movers and car travel times (i.e. column 5 in Table 4), *(iii)* the fixed-effects linear probability model based on movers and car travel times (i.e. column 9 in Table 4), *(iv)* the fixed-effects linear probability model based on movers and public transport travel times (i.e. column 10 in Table 4), and *(v)* the fixed-effects linear probability model based on stayers in the canton Zurich and public transport travel times (i.e. column 7 in Table 5). Overall, the graphs illustrate that the effect of distance is highly localized; the probability of forming a link is more than twice as large for direct neighbors than for people living 15 minutes apart. This probability continues to fall quickly up to a distance of 30 minutes, beyond which the negative effect of travel time flattens out. The convexity is strongest in the pooled OLS models, which suggest that part of the effect can be attributed to

[15] Besides, the estimations in columns (5) to (7) of Table 5 are restricted to the canton of Zurich (for callers) plus neighboring cantons (for callees) as this area experienced the most pronounced changes in public travel times. This further complicates direct comparisons.

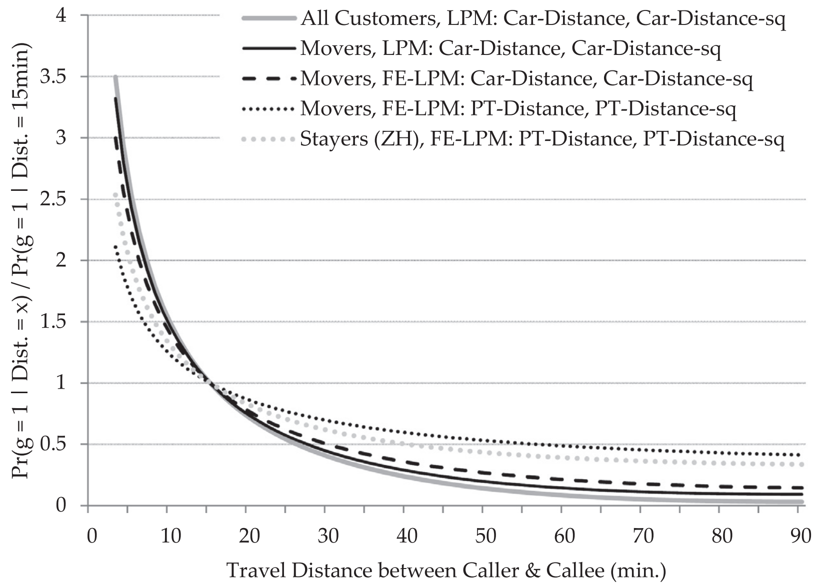


Fig. 3. Predicted Probability to Form a Link, Base Distance=15min Travel Time. *Notes:* All customers, LPM, car travel time: Table 4, column (2); Movers, LPM, car travel time: Table 4, column (5); movers, FE-LPM, car travel time: Table 4, column (9); movers, FE-LPM, public travel time: Table 4, column (10); stayers (ZH), FE-LPM, public transport travel time: Table 5, column (7). Models with controls are evaluated at the following values: Same Workplace=0, Common Contacts=0, Number of unique contacts=mean, Same Gender=1, Same Age=1, Age Difference=0, FE=0.

sorting. The distance effect in the public transport models fades out at somewhat higher travel times compared to the specifications that use car routing data. This is the pattern that one would expect, since public transport travel times are consistently larger than those for road travel.

We conclude from these findings that distance is highly detrimental to social interactions via mobile phones. Hence, this first set of tests unambiguously suggest that meeting face-to-face and phone communication are complementary. Next, we examine whether distance costs lead to differences in phone usage patterns across urban and rural areas.

5. Regional differences in density and phone usage

5.1. Empirical model

Each customer has a place of residence, r, which is assigned on the postcode level. We estimate the effect of location characteristics on individual-level phone usage statistics, namely total call frequency, total call duration, and number of unique contacts as well as the corresponding variables computed only for peers in the same catchment area. This yields six different phone call statistics – three for all calls and three for local calls within a 15 minutes perimeter around the billing address – which we log-transform and generically denote by C_{it}. Location-specific covariates at the place of residence are subsumed in vector L_r. Hence, we specify the benchmark model as

$$C_{ir,t} = \mu_i + L'_{r,t}\beta + X'_{ir,t}\gamma + \lambda_t + \lambda^l_r + \epsilon_{ir,t}, \tag{3}$$

where $X_{ir,t}$ is a vector of individual characteristics, λ_t stands for month fixed effects, and λ^l_r denotes language region fixed effects.

The location vector, $L_{r,t}$, includes indicators for EUROSTAT's harmonized definition of functional urban areas which distinguish between the urban core, the hinterland and peripheral regions. Alternatively, we measure local density using the number of private mobile phone customers within 15 minutes travel time from the respective place of residence. Unlike municipal population statistics this measure has the favorable feature that it is independent from administrative boundaries.

The vector of individual controls, $X_{ir,t}$, covers the customers' language (i.e. German, French, Italian, English), age (i.e. brackets spanning 10 years) and gender which are included as dummy variables in the pooled OLS-specifications without individual fixed effects. Beyond that, $X_{ir,t}$ also comprises two individual-location specific measures that can be combined with individual fixed effects: A dummy that captures whether an individual belongs to the local language majority, and a continuous measure for the commuting distance, which we infer based on the individual's billing address and the location of the transmitting antenna that the individual uses most frequently during business hours.

The issue of individual sorting on unobservables across locations is addressed by the fixed effect μ_i. Including μ_i requires time-changes in the location characteristics which stem in our benchmark models from movers. We consider those individuals as movers who shifted their residence by at least 30 minutes driving time such that they experience a substantial change in neighborhoods. This also ensures that the 15 minutes catchment areas used for calculating the local phone call statistics do not overlap. Beside focusing on long-range movers, we exclude a five months window around the moving date so that our estimates are not confounded by extraordinary calling behavior directly associated with the moving process. In this respect, Fig. A.3 reveals that the average number of unique phone contacts gradually increases three months prior to relocation, and then converges back to the pre-moving period within two months. By cutting out this 5 months window, we further allow the movers to adopt to their new environment.[16] If the most sociable individuals systematically sort into high-density places, specifications without individual fixed effect would yield upwardly biased estimates of the density externality. To infer the role of sorting we compare the fixed effects estimates to the pooled OLS specifications. In particular, we build on the following assumption: Unobservable personal characteristics that drive both the location choice and calling behavior are persistent during the twelve months we observe, and hence get absorbed by the individual fixed effect that we include when studying movers.

5.2. Main results

We now present the results on the second and third set of tests that assess the complementarity between mobile phone and face-to-face interactions. The higher population density (and hence lower distance costs)

[16] Excluding more post-move months tends to increase the effects associated with density, but also decreases the precision of the estimates.

Table 6
Regional Differences in Phone Usage Measures of Individuals.

a. Pooled OLS Models	Frequency			Duration			# Unique Contacts		
for All Customers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
City	0.120***			0.173***			0.022***		
	(0.001)			(0.001)			(0.001)		
Ln(Pop. Density)		0.028***			0.050***			-0.008***	
		(0.000)			(0.000)			(0.000)	
Population density									
75–90 percentile			0.031***			0.060***			-0.021***
			(0.001)			(0.001)			(0.001)
90–95 percentile			0.074***			0.124***			-0.004***
			(0.001)			(0.002)			(0.001)
95–99 percentile			0.143***			0.216***			0.018***
			(0.001)			(0.001)			(0.001)
$\geq$ *99 percentile*			0.192***			0.264***			0.046***
			(0.002)			(0.002)			(0.001)
Adj. R^2	0.061	0.059	0.062	0.031	0.030	0.032	0.067	0.067	0.068
Observations	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679
b. Fixed Effects Models	Frequency			Duration			# Unique Contacts		
for Sample of Movers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
City	0.041***			0.104***			0.011		
	(0.012)			(0.019)			(0.010)		
Ln(Pop. Density)		0.003			0.013*			-0.003	
		(0.004)			(0.006)			(0.003)	
Population density									
75–90 percentile			0.002			0.013			-0.011
			(0.014)			(0.021)			(0.012)
90–95 percentile			0.030^{+}			0.064*			-0.014
			(0.018)			(0.027)			(0.015)
95–99 percentile			0.042**			0.070**			0.019
			(0.016)			(0.024)			(0.013)
$\geq$ *99 percentile*			0.047^{+}			0.122**			-0.016
			(0.025)			(0.037)			(0.020)
Adj. R^2	0.771	0.771	0.771	0.684	0.684	0.684	0.724	0.724	0.724
Groups	16,679	16,679	16,679	16,679	16,679	16,679	16,679	16,679	16,679
Observations	90,011	90,011	90,011	90,011	90,011	90,011	90,011	90,011	90,011

Notes: We use monthly *outgoing calls* for June 2015–May 2016 to compute the logarithmized dependent variables. Note that population density is measured as the population within 15 minutes *road* travel time. The estimations in *Panel (a)* are based on *all customers* who used their phone every month at least once; they control for commuting distance, language of customer, dummy for belonging to a language minority, gender, age, language region, and month fixed effects. The estimations in *Panel (b)* are based on *movers* who used their phone every month at least once and changed their place of residence by at least 30 minutes driving time; they control for commuting distance, a dummy for belonging to a language minority, individual fixed effects, language region, and month fixed effects. Robust standard errors in parentheses. $+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

in cities raise the expected net benefits of interacting such that individuals increase the number of joint activities. At the same time one would expect that the mix of interactions shifts from electronic towards face-to-face interactions. If cities were to display a higher absolute level of phone interactions than low density places, this would imply that the first effect dominates and that the total number of interactions – via phone and face-to-face – is higher in cities.

Section 5.2.1 examines the relation between density and phone usage based on all calls. In Section 5.2.2, we re-compute the same phone usage measures for a subset of local calls, i.e. calls within a 15 minutes perimeter around the caller's place of residence. Both types of phone usage measures are examined for all available private mobile phone customers as well as a subset of customers who changed their place of residence between July 2015 and April 2016 by at least 30 minutes driving time. If systematic sorting of sociable individuals into cities exists, the pooled OLS models that include all customers overestimate the impact of density on phone usage intensity. Restricting the analysis to movers brings the advantage that we can back out time-constant unobservables.

5.2.1. Population density and phone usage: Call frequency, call duration, and number of contacts

We begin with the pooled OLS specifications that cover all private customers. Panel (a) of Table 6 shows the results for call frequency, call duration, and the unique number of contacts; the first column of each set of results contains the estimates for the discretized measure of urbanization, the second column examines continuous population density, and the third column assigns locations based on their population density into one of five groups, i.e. locations below the 75th-percentile, locations between the 75th- and 90th-percentile, locations between the 90th- and 95th-percentile, locations between the 95th- and 99th-percentile, and locations above the 99th-percentile.

According to the results in Panel (a), city residents have on average a 12 percent higher calling frequency, spend 17 percent more time on the phone, and call 2.2 percent more contacts than individuals living in the hinterland or periphery. The continuous population density measure is also positively correlated with calling frequency and calling duration, but negatively with the number of unique contacts. The latter finding is due to non-linearities in the lower ranges of population density, as the results in column (9) show. Including a squared-term also points toward a convex relation between population density and the average number of unique contacts (results not shown).

These positive correlations between density and phone usage are in line with similar patterns reported by Gaspar and Glaeser (1998) based on Japanese phone usage statistics and by Schläpfer et al. (2014) examining Portuguese mobile phone data. So far it remains unclear, however, whether these correlations can be interpreted as causal effects or are merely driven by the sorting of high sociability types into urban centers. If systematic sorting of sociable individuals into cities exists, the pooled OLS models overestimate the impact of density on phone usage. The estimation sample in Panel (b) is therefore restricted to individu-

Table 7
Regional Differences in *Local* Phone Usage Measures of Individuals Based on Calls within 15 Minutes Radius.

a. Pooled OLS Models	Frequency			Duration			# Unique Contacts		
for All Customers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
City	0.238***			0.392***			0.149***		
	(0.001)			(0.002)			(0.001)		
Ln(Pop. Density)		0.159***			0.335***			0.087***	
		(0.000)			(0.001)			(0.000)	
Population density									
75–90 percentile			0.182***			0.376***			0.087***
			(0.001)			(0.002)			(0.001)
90–95 percentile			0.235***			0.471***			0.122***
			(0.002)			(0.003)			(0.001)
95–99 percentile			0.329***			0.597***			0.197***
			(0.001)			(0.003)			(0.001)
$\geq$ *99 percentile*			0.413***			0.698***			0.272***
			(0.002)			(0.005)			(0.001)
Adj. R^2	0.037	0.051	0.042	0.018	0.033	0.022	0.051	0.059	0.054
Observations	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679	9,353,679
b. Fixed Effects Models	Frequency			Duration			# Unique Contacts		
for Sample of Movers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
City	0.066**			0.088			0.075***		
	(0.022)			(0.061)			(0.015)		
Ln(Pop. Density)		0.112***			0.286***			0.087***	
		(0.007)			(0.020)			(0.005)	
Population density									
75–90 percentile			0.100***			0.178**			0.046**
			(0.028)			(0.067)			(0.018)
90–95 percentile			0.079*			0.162^{+}			0.050*
			(0.031)			(0.089)			(0.022)
95–99 percentile			0.136***			0.253**			0.149***
			(0.030)			(0.079)			(0.020)
$\geq$ *99 percentile*			0.152***			0.342*			0.145***
			(0.043)			(0.121)			(0.031)
Adj. R^2	0.685	0.686	0.685	0.602	0.603	0.602	0.594	0.596	0.594
Groups	16,679	16,679	16,679	16,679	16,679	16,679	16,679	16,679	16,679
Observations	90,011	90,011	90,011	90,011	90,011	90,011	90,011	90,011	90,011

Notes: We use monthly *outgoing calls* for June 2015–May 2016 to compute the logarithmized dependent variables. Note that population density is measured as the population within 15 minutes *road* travel time. The estimations in *Panel (a)* are based on *all customers* who used their phone every month at least once; they control for commuting distance, language of customer, dummy for belonging to a language minority, gender, age, language region, and month fixed effects. The estimations in *Panel (b)* are based on *movers* who used their phone every month at least once and changed their place of residence by at least 30 minutes driving time; they control for commuting distance, a dummy for belonging to a language minority, individual fixed effects, language region, and month fixed effects. Robust standard errors in parentheses. $+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

als who changed their place of residence between July 2015 and April 2016. This allows us to include individual fixed effects that eliminate time-persistent characteristics.

The fixed effects estimates shown in Panel (b) of Table 6 substantially drop in magnitude compared to those obtained in the pooled OLS models: The results suggest that city residents spend 10 percent more time on the phone and have a 4 percent higher calling frequency than people living in the periphery or hinterland. Sorting evidently explains part of the pattern uncovered in the pooled OLS models but absorbing all time-constant individual characteristics does not eliminate the positive correlation between cities and phone usage measured via call frequency and call duration. Using the discretized density measures further shows that those locations with the highest population density make the most and longest phone calls: Compared to places below the 75-percentile in terms of density, those living in the 10 percent densest places make between 3 and 4.7 percent more phone calls (t-values: 1.7–2.6), and their calls last in sum 6.4 to 12.2 percent longer (t-values: 2.4–3.3). With regard to the total number of unique contacts things look different: the positive correlation between cities and the number of unique contacts completely disappears once spatial sorting is taken into account; the estimates fluctuate around zero and do not pass conventional thresholds of statistical significance (t-values: 0.8–1.5).

We interpret the latter finding as indication of a quality-quantity trade-off: Maintaining social contacts consumes time which acts as a direct constraint for the number of contacts. While the absolute number of contacts may not increase with population density the intensity of interactions does and the selection of contacts becomes more localized.[17] This is also consistent with results on social interactions from German survey data (see Burley, 2015) and fits well with the so-called *social brain hypothesis* claiming that the upper limit of group sizes is set by purely cognitive constraints (c.f. Dunbar,1992).

Overall, these findings lend strong support to the hypothesis that dense urban areas not only facilitate face-to-face interactions but also increase mobile phone usage suggesting that the two modes of social interactions are complementary. Between 40 and 70 percent of the higher phone usage in cities can be explained by sorting of sociable types to dense areas, as a comparison of the results in Panel (a) and (b) reveal. Yet, a robust effect of population density on call frequency and call duration remain, even when we account for time constant individual characteristics.

[17] In a working paper version we provide evidence for higher matching quality in cities which supports this interpretation (see Büchel and Ehrlich, 2017).

Table 8
Changes in Phone Usage Triggered by Shifts in Public Transport Accessibility, FE-Model.

a. Phone Usage Measures	Frequency		Duration		# Unique Contacts	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Pop. Density)	-0.003	-0.035*	-0.003	-0.016	-0.003	-0.018
	(0.002)	(0.017)	(0.003)	(0.021)	(0.002)	(0.011)
Ln(Pop. Density)^2		0.002*		0.001		0.001
		(0.001)		(0.001)		(0.001)
Adj. R^2	0.768	0.768	0.684	0.684	0.782	0.733
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,181,552	4,181,552	4,181,552	4,181,552	4,181,552	4,181,552
b. Local Phone Usage Measures	Frequency		Duration		# Unique Contacts	
i.e. within 15min. Radius	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Pop. Density)	0.168***	0.013	0.401***	0.133	0.130***	0.063*
	(0.009)	(0.041)	(0.019)	(0.092)	(0.006)	(0.030)
Ln(Pop. Density)^2		0.010***		0.017**		0.004*
		(0.003)		(0.006)		(0.002)
Adj. R^2	0.756	0.756	0.644	0.644	0.693	0.693
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,181,552	4,181,552	4,181,552	4,181,552	4,181,552	4,181,552

Notes: We use monthly *outgoing calls* for June 2015–August 2015 and March 2016–May 2016 to compute the logarithmized dependent variables such that we exclude the months close to the change in public transportation. We drop observations in the canton of Ticino as these were affected by an infrastructure change not recorded in our travel time data. Note that population density is measured as the population within 40 minutes *public transportation* travel time. The *sample* covers customers who used their phone every month at least once and did not change residence. *Further controls* include commuting distance and a dummy for belonging to language minority. Standard errors clustered by postcode in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

5.2.2. Population density and local phone usage: Local call frequency, local call duration, and number of local contacts

In this section we repeat the same analysis but recalculate the phone usage measures for calls within a 15 minutes perimeter around the caller's place of residence. City residents have an innately larger number of potential contacts within their catchment area. Thus, complementarity between face-to-face interactions and phone calls is especially bound to be reflected in a more localized network, that is a larger number of unique phone contacts, a higher frequency of calls, and a longer call duration within one's neighborhood.

The reported correlations from pooled OLS models in Panel (a) of Table 7 are statistically highly significant and show that people living in dense areas make more calls within their catchment area, these calls take more time, and are directed to a larger number of unique contacts. The measured differences are quite substantial: City residents make on average 24 percent more calls to individuals within a 15 minutes radius, the sum of these calls takes 39 percent longer, and city residents reach 15 percent more different contacts within their catchment area. Columns (3), (4) and (5) confirm that the frequency and duration of local calls as well as the number of unique local contacts is highest among those individuals that live in the 25 percent most densely populated areas.

Panel (b) of Table 7 restricts the analysis to movers which allows us to absorb all individual specific time constant heterogeneity. These estimates also back the hypothesis that face-to-face and phone interactions are complementary: The effect of density on phone usage for local calls is consistently positive and 17 out of 18 coefficients are significant at the 10 percent level or higher. Again, the fixed effect estimates drop in magnitude compared to those in the pooled OLS models: A comparison of the estimates in Panel (a) and (b) suggest that roughly 50 to 70 percent of the difference in local phone usage can be explained by sorting of more sociable types into urban areas. Nonetheless the remaining impact of density is considerable: On average, doubling the population density increases local calls by 11 percent, their duration by 28 percent, and the number of unique contacts rises by 9 percent.

5.3. Robustness analysis

So far our analysis shows a coherent relation between population density and phone usage: Individuals in cities use their phones more intensely than individuals in rural areas even when sorting is accounted for. This gap shows in phone usage measures based on all calls and – even more pronounced – for local calls within a 15 minutes travel time perimeter. Both patterns support the hypothesis that mobile phone calls complement (not substitute) face-to-face interactions.

In this section, we examine the sensitivity of these results along a number of dimensions. First, we address the concern, that preference changes driving location choice and mobile phone usage may coincide, which would invalidate the identification assumption imposed in the benchmark analysis based on movers. Second, we instrument local population density to account for unobserved location characteristics that might be correlated with both population density and phone usage. Third, we analyze whether the main conclusions also hold for phone usage measures that incorporate text messages. Finally, we explore several dimensions of effect heterogeneity.

5.3.1. Adjustments in the phone calling behavior of stayers following a revision of public transport schedules

An issue of our analysis in sections 5.2.1 and 5.2.2 may be that the factors driving the moving decision are potentially correlated with phone usage behavior. For instance, if one develops a taste for diverse cultural events and vibrant socializing and hence decides to move to a city, the simultaneous increase in phone usage and local population density may be wrongly interpreted as causal relation. Similarly, systematic life cycle patterns that drive location decisions and mobile phone usage may bring about correlations between local density and phone usage measures that are mistakenly interpreted as causal effect.

In the following, we look at catchment areas defined by public transportation travel time rather than road travel time. We measure population density by the number of private mobile phone customers living within a 40 minutes public transport travel time perimeter from the

Table 9
Regional Differences in Phone Usage of Movers, IV-FE Model.

a. Phone Usage Measures	Frequency		Duration		# Unique Contacts	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Pop. Density)	0.008		0.033***		-0.000	
	(0.007)		(0.010)		(0.006)	
City		0.036		0.146***		-0.001
		(0.029)		(0.044)		(0.024)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Lang. Region & Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Groups	15,898	15,899	15,898	15,899	15,898	15,899
Observations	84,423	84,430	84,423	84,430	84,423	84,430
b. *Local*Phone Usage Measures,	Frequency		Duration		# Unique Contacts	
i.e. within 15min. Radius	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Pop. Density)	0.089***		0.188***		0.053***	
	(0.012)		(0.033)		(0.008)	
City		0.383***		0.815***		0.228***
		(0.051)		(0.143)		(0.036)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Lang. Region & Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Groups	15,898	15,899	15,898	15,899	15,898	15,899
Observations	84,423	84,430	84,423	84,430	84,423	84,430

Notes: We use monthly *outgoing calls* for June 2015–May 2016 to compute the logarithmized dependent variables. Current population density is measured as the population within 15 minutes road travel time of a postcode. The *instrument in the first-stage* regression is log population density at the municipality level in 1850, for which we obtain coefficients of 0.767 (t-value=227.3) in columns (1), (3), (5) and 0.176 (t-value=128.6) in columns (2), (4), (6). The *sample* consists of movers who used their phone every month at least once. *Further controls* include commuting distance, dummy for belonging to language minority. Robust standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

respective place of residence. We have chosen the 40 minutes threshold, because it yields population densities of catchment areas that are comparable to the ones obtained for 15 minutes road travel time.[18] Since this measure of local population density is a function of public transport accessibility, it is affected by the revision of public transport schedules described in Section 4.2.2. The distribution of changes in local population density includes positive and negative values with the 10th-percentile change being a 8 percent decrease in density, the median change equaling plus 1 percent, and the 90th-percentile change amounting to an 37 percent increase. Analogously, we now compute the local measures of calling frequency, duration of calls, and number of unique contacts based on calls directed to people living in catchment areas of 40 minutes public transport travel time.

Using this new set of variables, we re-estimate the specifications in Eq. (3) and now identify the impact of population density through changes triggered by the revision of the public transport schedule. Importantly, a comparison of (pre-shock) characteristics between those individuals that reside in postcodes with below median changes in effective population density with those individuals that live in postcodes with above median changes in effective population density, show that the two groups are virtually identical in terms of (pre-shock) calling behavior and sociodemographics (see Table A.6 in the appendix). The results obtained from examining this natural experiment are summarized in Table 8 where panel (a) displays the global phone usage statistics and panel (b) shows the local phone usage measures. Note that in contrast to the estimates in Section 5.2.1 and 5.2.2 we do not use the binary city indicator because the change in public transport times has no consequences for the assignment of postcodes to urban versus non-urban areas.

Panel (a) shows that the linear specifications yields insignificant effects of population density on all three measures of phone usage. Except for calling duration this was also true in the fixed effect specifications based on movers in Table 6 which was explained by a convex relationship. In columns (2), (4), and (6) we add again a quadratic term. The marginal effect of population density on calling frequency is positive and significant for catchment areas of above ca. 6300 inhabitants which is well below the average size of catchment areas. The estimated coefficients confirm a convex relationship for calling frequency while for calling duration we do not find significant effects.

The analysis of local phone usage patterns shows a much clearer picture. The estimated coefficients of population density in panel (b) of Table 8 are highly significant and similar in magnitude as the estimates obtained in the movers analysis: According to the estimates in panel (b) of Table 8, a ten percent increase in population density causes a 1.7 percent increase in local calling frequency, a 4 percent higher calling duration within the catchment area, and 1.3 percent more unique local contacts; the corresponding values in the movers analysis, i.e. panel (b) of Table 7, were 1.1 percent, 2.9 percent, and 0.9 percent.

This set of evidence is based on a relatively small but arguably exogenous change in effective distances and corresponding catchment areas, affecting only users of public infrastructure. It supports the findings obtained for movers and again shows that higher population density leads to more local calls and – albeit to a lesser extent – a higher general level of phone usage as reflected in the frequency of calls.

5.3.2. Accounting for unobserved place characteristics

Another source of bias may be unobserved location characteristics. If such unobserved location features affect both phone usage and population density, we would obtain biased estimates. We address this concern by adopting an IV-strategy to instrument current population density with historical population counts (see Ciccone and Hall, 1996; Combes et al., 2010). Hence, we aim to exploit exogenous variation in population density that has been determined by historical factors. The idea is that historical population counts (year 1850) are unlikely to have a direct effect on phone usage measured today. Yet, they are a strong predictor of population density today as is evident from the first-stage regressions.

In Table 9 we re-estimate the fixed-effect specifications based on movers in a two-stage least square framework and using historical pop-

[18] An average catchment area counts 11,950 customers when a 40 minutes public transport travel time perimeter is used, while it includes 14,683 customers for a 15 minutes road travel time perimeter.

Table 10
Regional Differences in Phone Usage (Calls & Text Messages) of Movers, FE-Model.

	All Outgoing Activity				*Local* Activity within 15 min. Radius			
	Frequency		# Unique Contacts		Frequency		# Unique Contacts	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
City	0.023^{+}		0.006		0.140***		0.152***	
	(0.013)		(0.009)		(0.025)		(0.013)	
Ln(Pop. Density)		-0.003		-0.004		0.144***		0.118***
		(0.008)		(0.003)		(0.008)		(0.004)
Adj. R^2	0.747	0.747	0.740	0.740	0.671	0.672	0.623	0.626
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lang. Reg. & Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	90,011	90,011	90,011	90,011	90,011	90,011	90,011	90,011

Notes: We use monthly *outgoing text messages and calls* for June 2015–May 2016 to compute the logarithmized dependent variables. Note that population density is measured as the population within 15 minutes *road* travel time. The *sample* consists of movers who used their phone every month at least once. *Further controls* include commuting distance, dummy for belonging to language minority. Robust standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

ulation counts as instrument for current population density.[19] The estimates generally confirm the results of sections 5.2.1 and 5.2.2. The magnitude of the effects on calling frequency and duration is very similar to the one estimated in Table 6 but shows higher standard errors for calling frequency. The effects on the number of unique contacts remain insignificant as throughout our analysis. With regard to the local phone usage measures, the estimates prove more robust: For all three outcomes the estimated effects are significant, positive, and have a similar magnitude as the corresponding estimates obtained in the benchmark analysis.

5.3.3. *Phone usage measures incorporating text messages*

One issue may be that our results are not driven by the actual volume of interactions but rather by different preferences for voice calls and text messages. Our main reason to focus on calls is the growing popularity of messenger apps (like WhatsApp) during our period of study. These messenger apps are a natural substitute for traditional text messages (like SMS and MMS) but less for voice calls. As Table A.1 in the appendix shows, this reflects in steep decline of text messages via our provider's network, while the volume of voice calls remained almost constant. Hence, we are more concerned about systematically different adoption rates of messenger apps in cities versus rural areas than about different preferences for texting versus calling across space. It seems more likely that people in urban areas adopt the new technology more quickly than rural residents, which – if anything – would lead to a downward bias in our estimates of density on phone usage volumes.[20]

In Table 10 we examine the robustness of our benchmark analysis with movers by including text messages in our phone usage measures. The effect of cities becomes somewhat weaker for total phone usage, but overall the main conclusions remain robust. Phones are used more frequently in cities than in rural areas, and this difference gets pronounced if we focus on a local calls and messages within a 15 minutes catchment area.[21]

5.3.4. *Effect heterogeneity*

The identification strategy based on movers may raise the concern that movers differ systematically from the rest of the population. As observed in other studies adopting a similar estimation strategy (e.g. D'Costa and Overman, 2014), movers in our data are on average younger than non-movers (see Table A.4 in the appendix). We now examine whether the impact of population density on local calling behavior is heterogeneous across subgroups of the population. Table B.5 in the appendix focuses on the local phone usage measures of as these turned out to display the most pronounced effects.

It appears from this analysis that the effects are somewhat asymmetric with respect to the moving direction: In particular, the estimated coefficients are significant for those moving out of a city, while they drop in magnitude and become insignificant for those that move into the city. Yet, the discretized city indicator conceal some of the action, as the asymmetry is considerably less pronounced when using the continuous measure of population density.

In terms of age, we find that the effects get muted among the youngest age group, while no clear pattern emerges for those aged between 35 to 44 and between 45 to 64. Considering that eight out of nine coefficients remain significant when estimated separately for the three age groups, the overall pattern appears very robust in this dimension.

With respect to gender, we find that the local phone usage of men reacts stronger to changes in density than the calling behavior of women. Yet the qualitative insights remain unaffected: Both men and women use their phone more extensively for calls within a 15 minutes travel time radius when they move to a more densely populated area.

Overall, the data unveils some heterogeneity across different sociodemographic groups, but the main conclusions are not substantially altered by this robustness exercise: Population density increases the frequency and duration of local calls which supports the hypothesis that face-to-face interactions and mobile phone calls are complements.

6. Conclusions

This study analyzes social interactions at a very granular level and demonstrates that the probability of social interactions via mobile phones decline as the distance between individuals increases. The estimated distance elasticity converges to zero if travel time exceeds about

[19] The instrument in the first-stage regression yields coefficients on historical log population density of 0.77 (t-value=227.3) in columns (1), (3), and (5) and 0.18 (t-value=128.6) in columns (2), (4), and (6). In a previous version we have also used local soil quality as an instrument which confirms the results.

[20] We also decompose messenger usage along gender and language region based on a survey conducted by *comparis.ch* in 2014. It shows that messenger apps are more often used among men than women and are more widespread in French-speaking than German-speaking regions. The same ranking unfolds for usage intensity of mobile phones. This indicates that the two media are complements not substitutes.

[21] In Section B.3 in the appendix, we discuss another robustness analysis where we vary the applied pre-filtering of the phone data. Table B.3 shows that including both incoming and outgoing calls does not alter the conclusions from our benchmark analysis.

100 minutes, and among city residents this threshold is reached at even smaller distances. The fact that almost 50 percent of phone contacts are established between people living within a 5 kilometer radius, also illustrates how localized social interactions via phone are.

Phone usage, as measured by calling frequency and duration, significantly increases with population density, and this effect is even more pronounced if one focuses on local calls. While about 50 percent of this effect is driven by sorting of heterogeneous individuals across locations, we show that significant and sizable differences in phone usage remain when sorting is accounted for. We find an elasticity of local calling frequency with regard to population density of 0.1 to 0.2 and an elasticity of calling duration with regard to population density of 0.3 to 0.4. While the number of unique contacts is unaffected by density when sorting is accounted for, our estimates suggest that the number of local contacts increases by about 10% when population density is doubled.

All our evidence indicates a complementary relationship between face-to-face and phone interactions. Given this complementarity, mobile phone interactions may provide a valuable proxy for social networks in order to study the effects of density on further dimensions of network structure which have been addressed in urban economic theory (e.g. Berliant et al., 2006; Helsley and Zenou, 2014; Sato and Zenou, 2015). From a policy perspective, our results provide micro-level evidence for the positive externalities of densely populated areas, which should be taken into account, for example, in the design of zoning policies, or the pricing of mobility.

Appendix A. Data

A1. Overall phone usage statistics

Table A.1 displays monthly phone activity and call duration statistics of private customers subdivided into device and message type, i.e. mobile phone calls, text messages sent from mobile phones and landline calls. We restrict our analysis to mobile phones and then filter the data as motivated in Section 3. In particular we restrict the analysis to the first 28 days of a month, *keeping* (*i.a.*) calls between mobile phones, (*i.b.*) customers that registered only one mobile phone, (*ii.*) outgoing calls, (*iii.*) calls with a duration of more than 10 seconds, (*iv.*) mobile phones with a monthly call duration between 1 minute and 56 hours. The filtered data comprises about 40% of private mobile phones calls in the data representing 60% of the total call duration; the filtering skews the sample towards relatively long-lasting calls as very short calls are deleted from the data set in step (*iii.*).

A2. Calling behavior and sociodemographic characteristics

Table A.2 shows summary statistics of calling behavior and the main demographics by age, gender, and type of residence.

Fig. A.1 plots various phone usage statistics. It illustrates, that the distribution of call frequency (unfiltered), call duration (unfiltered) and the number of unique contacts (i.e. degree centrality) is markedly right-skewed: For instance, the average degree in our monthly data is about 10, with the vast majority having a degree below 20 and some hub-

Table A.1
Call Duration (in Mio. Minutes) between June 2015 to May 2016.

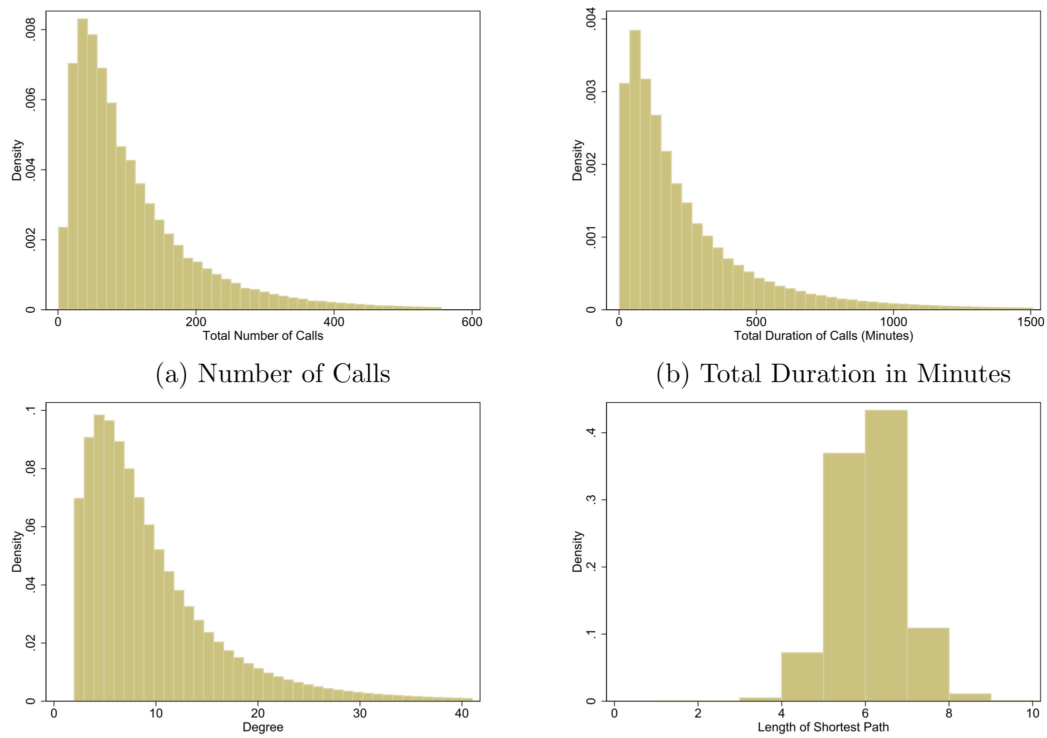
	Phone Activity (in Mio.)					Call Duration (in Mio. Minutes)			
	MP-Calls	SMS	Landline	Total	*Filtered*	MP-Calls	Landline	Total	*Filtered*
Jun. 2015	166.3	90.9	64.3	**321.6**	*66.0*	351.2	296.2	**647.4**	*222.4*
Jul. 2015	157.3	91.9	57.8	**307.0**	*62.0*	324.8	271.1	**595.9**	*202.2*
Aug. 2015	153.6	89.0	59.7	**302.3**	*60.3*	337.0	283.6	**620.6**	*211.3*
Sep. 2015	153.8	85.2	61.9	**300.9**	*61.6*	343.0	294.2	**637.2**	*216.9*
Oct. 2015	133.6	76.3	59.9	**269.8**	*53.7*	307.5	284.8	**592.3**	*192.6*
Nov. 2015	138.1	77.7	62.1	**277.9**	*56.5*	333.1	298.5	**631.6**	*208.7*
Dec. 2015	154.1	79.1	61.6	**294.8**	*62.0*	347.4	298.1	**645.5**	*218.5*
Jan. 2016	155.7	78.5	62.0	**296.2**	*61.0*	376.0	312.4	**688.4**	*235.5*
Feb. 2016	167.6	77.5	60.6	**305.7**	*66.3*	393.3	299.6	**692.9**	*246.7*
Mar. 2016	163.3	74.9	58.6	**296.8**	*65.4*	378.1	286.8	**664.9**	*240.3*
Apr. 2016	164.2	70.7	59.9	**294.8**	*65.7*	378.8	286.1	**664.9**	*241.1*
Mai 2016	161.1	68.6	55.9	**285.7**	*64.9*	353.5	264.6	**618.1**	*228.3*

Notes: These figures base on phone usage statistics of private customers.

Table A.2
Calling Behaviour and Sociodemographics by Age, Gender and Region.

Sample means, sd in brackets	Total	Age 15–24	Age 25–44	Female	City
Number of Calls	32.332	34.868	34.338	28.800	35.900
	(37.286)	(38.230)	(39.115)	(31.826)	(41.919)
Number of *Local* Calls	21.641	24.927	22.103	20.193	27.011
	(32.250)	(34.722)	(33.138)	(30.145)	(39.805)
Duration of Calls	117.176	114.917	129.890	121.449	131.142
	(167.929)	(173.437)	(173.238)	(170.920)	(179.299)
Duration of *Local* Calls	62.187	64.866	67.466	65.388	81.434
	(127.124)	(135.980)	(130.629)	(133.181)	(157.022)
Number of Unique Contacts	9.214	9.311	9.588	7.666	9.360
	(7.874)	(7.024)	(8.418)	(5.795)	(8.029)
Number of Unique *Local* Contacts	7.074	8.182	6.747	6.027	8.019
	(7.221)	(7.403)	(7.323)	(5.656)	(7.940)
Age	34.964	20	33.718	36.186	35.080
	(13.561)	(-)	(4.833)	(13.900)	(13.042)
Female	0.522	0.485	0.506	1	0.523
	(-)	(-)	(-)	(-)	(-)

Notes: Local refers to the subset of calls within a radius of 15 minutes around an agent's residence. The table is based on the subsample of customers with phone activity in all 12 months, which we also use in the main analysis. Further filters as described in Section 3.



(a) Number of Calls (b) Total Duration in Minutes (c) Number of Unique Contacts (d) Dyad Distances

Fig. A.1. Phone Usage Statistics & Network Characteristics for June 2015 *Notes:* Subfigure (a) and (b) are based on unfiltered data, while the data for subfigure (c) and (d) was filtered as described in Section 3. *(d) Dyad Distances:* Length of shortest paths connecting 100 randomly selected agents with every other private mobile phone user in the data.

agents reaching network sizes of 100 links or more. As reported in other studies on social networks, the probability distribution of degree centrality is well fitted ($R^2 = 0.92$) by a power-distribution, $P(D) = cD^{-\varphi}$, with parameter estimates of $\hat{\varphi} = 3.86$ and $\hat{c} = 5.96$.

To gain insights into the diameter and the average path length, we randomly select 100 individuals and calculate the length of the shortest paths connecting every other private mobile phone user in the data. The mean path length in the sample is 5.6, with the longest path having a length of 12; the histogram plotted in Fig. A.1d reveals that 88 percent of dyads are separated by 6 or fewer links. This fits strikingly well with the "small-world"-hypothesis first formulated by Milgram (1967) and the early empirical evidence based on a chain letter experiment conducted by Travers and Milgram (1969).

A3. Representativeness of customer data

Table A.3 displays the correlation coefficients of population figures from the census data and the customers numbers from our data by age group, language group, and gender. It is evident that our data is highly representative for the Swiss population at the local level. This holds even true when we study specific subgroups of the population as the correlation coefficients are always well above 0.9 except for Italian speaking part of Switzerland (Ticino). In Ticino, which represents only about 5 percent of Swiss municipalities, we still observe a correlation coefficient of about 0.9 but other phone providers seem to be relatively strong for the age group 30 where the correlation coefficients is only 0.765.

Table A.3
Correlation between Census Population and Number of Customers at the Municipality Level.

	All	Male	Female	German	French	Italian
Age All	0.987	0.984	0.988	0.992	0.990	0.893
Age 20	0.945	0.946	0.944	0.960	0.946	0.916
Age 30	0.953	0.955	0.951	0.953	0.973	0.765
Age 40	0.968	0.963	0.971	0.983	0.993	0.875
Age 50	0.985	0.982	0.984	0.993	0.988	0.914
Age 60	0.990	0.988	0.987	0.994	0.984	0.922

Notes: These figures base on customer information of active phones during June 2015 and the most recent census conducted by the Federal Statistical Office in 2010.

A4. Representativeness of movers

One concern maybe that movers are systematically different from non-movers. Table A.4 compares phone usage statistics and sociodemographics between movers and non-movers. While movers are considerably younger than non-movers (∼ 4.2 years, ∼ 1/3 SD), their phone usage behavior only differs marginally. In terms of gender and language, movers are practically identical to non-movers.

A5. Changes in public transport accessibility due to a major revision of the federal railway timetable

In Sections 4.2.2 and 5.3.1 we exploit changes in the federal railway timetable to infer the causal impact of distance and density on calling behavior. The new timetable was put into effect on 13 December 2015, splitting our sample of phone data – that spans from June 2015 to May 2016 – into 6.5 and 5.5 months periods. Notably, the planning of Switzerland's public transport schedules is considerably centralized; the Swiss Federal Railways company (SBB) holds a market share of around 80% in rail traffic so that local providers typically coordinate their services with the SBB timetable. For instance, Switzerland's largest city transport network in Zurich – the Zōrcher Verkehrsverbund (ZVV) – also revised its timetables on 13 December 2015 in order to synchronize their

Table A.4
Comparing Non-movers to Movers, Main Descriptive Statistics.

	Non-Movers		Movers	
	Mean	SD	Mean	SD
Monthly Phone Usage Statistics, June 2015 – May 2016 (pooled)				
Number of Calls	32.060	37.183	35.818	38.519
Number of *Local* Calls	21.777	32.295	20.361	38.519
Duration: Calls (Minutes)	115.376	166.764	140.340	180.687
Duration: *Local* Calls	62.128	127.008	64.553	131.598
Number of Unique Contacts	9.178	7.878	9.614	7.784
Number of Unique *Local* Contacts	7.168	7.256	6.009	6.743
Sociodemographics - Private Mobile Phones				
Age	35.273	13.734	31.091	10.632
Female	0.522	–	0.529	–
Language: German	0.680	–	0.703	–
Language: French	0.271	–	0.251	–
Language: Italian	0.043	–	0.039	–
Language: English	0.006	–	0.007	–

The table is based on the subsample of customers with phone activity in all 12 months, which we also use in the main analysis. It covers 797,053 stayers and 69,593 movers. Further filters as described in Section 3.

connections with SBB. This centralization facilitates reliable timetable queries from websites such as *search.ch*, and also brings about nationwide changes in public transport connections triggered by revisions in SBB's scheduling.

The change of timetable in December 2015 was the largest of its kind since 2004. It aimed at incorporating new regional and interregional connections affecting travel times both through longer/shorter journey times and through longer/shorter waiting times.[22] To calculate the changes in travel time between two places, we proceed as follows: *search.ch* kindly provided data on the quickest connections between all pairs of serviced public transport stops, i.e. about 26,000 $\times$ 26,000 different routes, including the frequency of available connections for a two hour window between 6am and 8am. The data covers four randomly chosen weekdays in 2015 (before the change of the timetable) and four randomly chosen weekdays in 2017 (after the change of the timetable). We build a cleaned and integrated file for 2015 and 2017, where we select the day with the shortest journey time; typically journey times do not vary across different days of the week unless construction or maintenance work causes temporary delays. As the data includes x-y-coordinates of each public transport stop, we can reliably assign them to a postcode/municipality; we then extract the quickest transport link for every postcode/municipality pair in 2015, including the stop-ids, journey time in minutes, and the number of available connections between 6am and 8am. Our final measure of public transport travel times incorporates both journey and waiting time, and is defined as

$$Public\ Transport\ Travel\ Time = \frac{120\ min.}{\#Available\ Connections} + Journey\ Time. \tag{A.1}$$

To obtain a comparable travel time matrix for 2017, we use the same selection of stops as in the 2015 matrix. Any changes between travel times in 2015 and 2017 can then be attributed to the change of timetable on 13 December 2015. Fig. A.2 plots the distribution of percentage changes in the travel time between 2015 and 2017, while summary statistics for public transport travel times in 2015 are shown in Table 2. The largest changes occurred around Zurich, which is why we estimate the models for Switzerland as well as a subsample consisting of Zurich and its neighboring cantons, namely Schaffhausen, Thurgau, St. Gallen, Schwyz, Zug, and Aargau. Note that we exclude postcodes from the Italian speaking canton of Ticino as the opening of a new north-

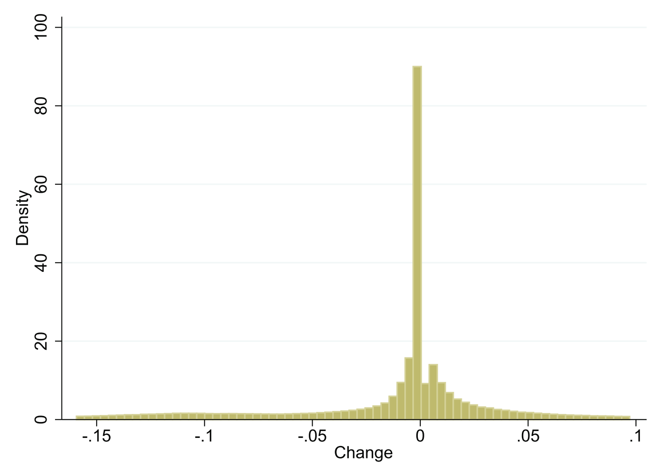


Fig. A.2. The Impact of the Revised Timetable on Travel Times between Postcode Pairs *Notes:* Illustrates the LN-differences for travel times before and after the change in the federal railway timetable on 13 December 2015; Mean: -0.012, Std. Dev.: 0.084, Share of Zeros: 0.256. Postcodes in the canton of Ticino are excluded, as these were affected by an infrastructure change not recorded in our travel time data.

south connection in December 2016 primarily affected public transport schedules in this part of Switzerland; these changes would be wrongly attributed to the re-scheduling in December 2015 that we exploit for our analysis.

The change in the federal railway timetable affected travel times for three quarters of postcode pairs; on average the modifications lowered travel times by 1.2% ranging from reductions of 15% up to increases of 10%. When we measure population density by the number of private mobile phone customers living within a 40 minutes public transport travel time perimeter from the respective place of residence, these changes also have implications for the effective population density. The distribution of changes in local population density includes positive and negative values with the 10th-percentile change being a 8 percent decrease in density, the median change equaling plus 1 percent, and the 90th-percentile change amounting to an 37 percent increase.

One may be concerned that people anticipated this major revision of public transport schedules and adopted their behavior already prior to its implementation. This seems unlikely, however. Surveying the archive of press releases by the SBB suggest that a draft of the new schedule was first communicated on 26 May 2015. At that time the final time tables

[22] Detailed summaries of all changes made in December 2015 can be found on the SBB's website, e.g. https://stories.sbb.ch/fahrplanwechsel-dezember-2015/2015/11/10/ or https://company.sbb.ch/de/medien/medienstelle/medienmitteilungen/detail.html/2015/11/1111-1.

Table A.5
Descriptive Statistic for Stayers Living in Postcodes with Above/Below Median Absolute Changes in Public Transport Travel Times.

	Below Median Change		Above Median Change	
	Mean	SD	Mean	SD
Monthly Phone Usage Statistics, June 2015 – November 2015 (pooled)				
Number of Calls	31.590	37.038	31.451	36.329
Number of *Local* Calls	21.538	32.184	21.217	31.164
Duration: Calls (Minutes)	115.173	166.875	112.281	163.350
Duration: *Local* Calls	61.993	127.054	60.221	123.530
Number of Unique Contacts	9.051	7.755	9.153	7.875
Number of Unique *Local* Contacts	7.079	7.186	7.177	7.251
Sociodemographics - Private Mobile Phones				
Age	35.233	13.662	35.239	13.748
Female	0.524	–	0.525	–
Language: German	0.748	–	0.660	–
Language: French	0.238	–	0.330	–
Language: Italian	0.007	–	0.005	–
Language: English	0.007	–	0.005	–
Share City Residents	0.242	–	0.184	–

The table is based on the subsample of customers with phone activity in all 12 months, which we also use in for the analysis presented in Table 5. It covers 394,434 stayers living in postcodes with below median changes in public transport accessibility and 359,280 stayers living in postcodes with above median changes in public transport accessibility. Postcodes in the canton of Ticino are excluded, as these were affected by an infrastructure change not recorded in our travel time data. Further filters as described in Section 3.

Table A.6
Descriptive Statistic for Stayers Living in Postcodes with Above/Below Median Changes in Population Density Due To Adjustments in Public Transport Accessibility.

	Below Median Change		Above Median Change	
	Mean	SD	Mean	SD
Monthly Phone Usage Statistics, June 2015 – November 2015 (pooled)				
Number of Calls	31.411	36.573	31.646	36.840
Number of *Local* Calls	21.627	31.851	21.123	31.537
Duration: Calls (Minutes)	113.186	165.021	114.453	165.411
Duration: *Local* Calls	61.8788	125.863	60.359	124.870
Number of Unique Contacts	9.058	7.770	9.145	7.858
Number of Unique *Local* Contacts	7.189	7.276	7.057	7.153
Sociodemographics - Private Mobile Phones				
Age	35.249	13.750	35.223	13.653
Female	0.525	–	0.523	–
Language: German	0.696	–	0.716	–
Language: French	0.291	–	0.273	–
Language: Italian	0.006	–	0.006	–
Language: English	0.007	–	0.005	–
Share City Residents	0.223	–	0.204	–

The table is based on the subsample of customers with phone activity in all 12 months, which we also use in for the analysis presented in Table 8. It covers 391,901 stayers living in postcodes with below median changes in public transport accessiblity and 361,813 stayers living in postcodes with above median changes in public transport accessibility. Postcodes in the canton of Ticino are excluded, as these were affected by an infrastructure change not recorded in our travel time data. Further filters as described in Section 3.

were not fully developed yet; as pointed out in the press release, the schedules of the local providers still needed to be harmonized with that of the federal railway company. The main announcement on the revision of time tables was issued on 11 November 2016, one month before the changes took effect.[23] Placebo checks in Table B.2 also dismiss the concern that people adapted their behavior in advance.

A6. Representativeness of individuals affected by change in public transport accessibility

[23] All press releases by the SBB can be found online, see https://company.sbb.ch/de/medien/medienstelle/medienmitteilungen.html (last access: 31.08.2019).

A7. Moving dynamic and number of unique contacts

Fig. A.3 plots the dynamics of the number of unique contacts (degree centrality) for movers (minimum distance of 30 min. driving time) around the moving month. The y-axis depicts the deviation in degree centrality relative to the first month at the new address, while the x-axis reflects the timeline in terms of relocation.

This event-study type of graph reveals that the average degree centrality of movers gradually increases three months prior to relocation, and then converges back to the pre-moving period within two months.

In our analysis on the calling behavior of movers (i.e. Tables 6, 7, 9, B.3, 10, B.5) we exclude a five months window around the moving date so that our estimates are not confounded by extraordinary calling behavior directly associated with the moving process. By cutting out this 5 months window, we further allow the movers to adopt to their new environment.

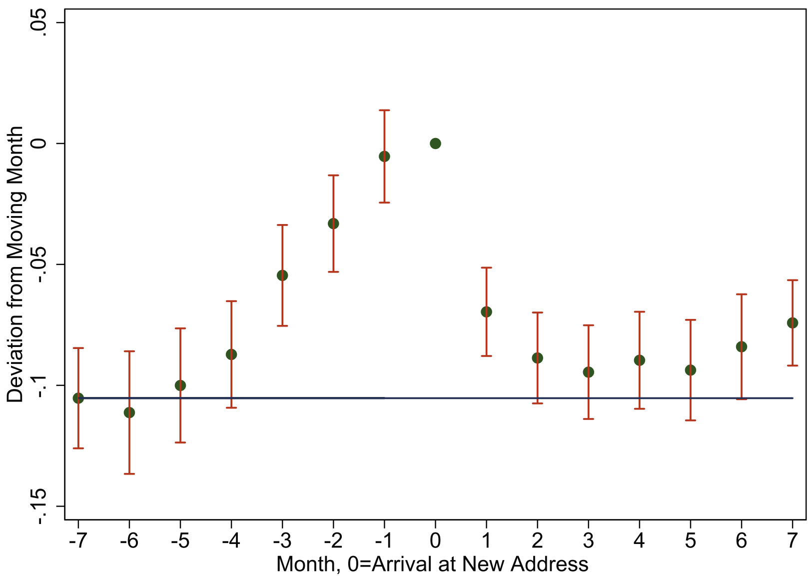


Fig. A.3. The Number of Unique Contacts Prior and After Moving
Notes: We regress the number of unique contacts (degree centrality) of movers on dummies for the months leading and following the moving date. The reference category is the moving date $t = 0$ which we define as the first month at the new location. The lines illustrates the 95 percent confidence bounds around the point estimates.

In an earlier draft of this paper, we step-by-step exclude periods around the moving date ($t = 0$), which we define as the first month at the new residence in order to test the robustness of the results. Excluding more post-move months tends to increase the effects associated with density, but also decreases the precision of the estimates (results not reported).

Appendix B. Additional results to assess robustness and heterogeneity

B1. Distance and social interactions: Logit specification and distance bins

We also estimate Logit models of link formation to accommodate for the binary dependent variable and check the robustness of these results. Since the incidental parameter problem can induce severe bias in the logit fixed effects estimates (e.g Lancaster, 2000), Table B.1 only shows results for the pooled logit model. Fig. B.1 plots the functional form together with the pooled OLS specification that substitutes travel time with a series of 5 minutes car travel distance bins.

Table B.1
Link Formation, Logit.

	(1)	(2)
Ln(Travel Time$_{ij,t}$)	-1.548***	-0.877***
	(0.001)	(0.049)
> 0 Common Contacts		7.363***
		(0.122)
> 1 Common Contacts		2.323***
		(0.352)
Const.	-6.746***	-12.951***
	(0.009)	(0.249)
Pseudo R^2	0.198	0.393
Further Controls	No	Yes
Month FE	Yes	Yes
Observations	49,172,284	49,172,284

Notes: The *sample* as in columns (1) & (2) of Table 4. *Further controls*: same workplace, same language, number of unique contacts of both agents, dummies for same gender and same decimal age-bracket, as well as the absolute age difference between agents i and j. Standard errors in parentheses. *** $p < 0.001$.

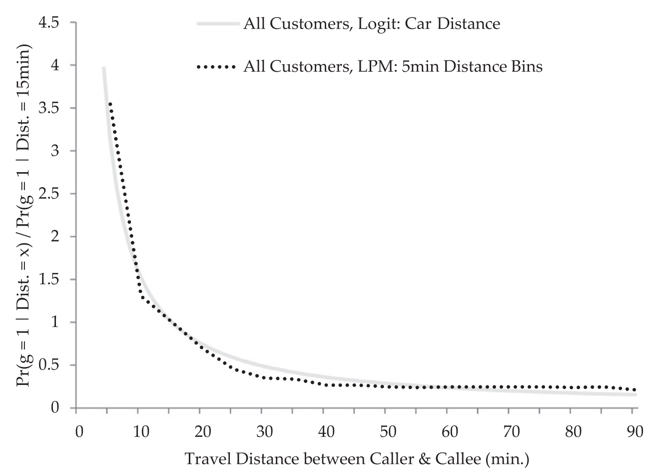


Fig. B.1. Predicted Probability to Form Link.

B2. Placebo check for changes in calling behavior following "Fake" revision of public transport schedules

We address the possibility that our specifications in Section 4.2.2 pick up trends in social interactions that prevailed already prior to the actual change in the public transportation timetable. To this end we perform a series of placebo checks where we set 'fake' changes in the timetable half a year prior to the actual change. In particular we use the actual change in the bilateral travel times between postcode pairs but artificially time them in September 2015 (instead of December 2015). We run these specifications for 20 samples following the sampling strategy described in Section 4.1. Table B.2 shows the corresponding results: None of the estimates yield a significant effect of the 'fake' travel time changes on link formation and the magnitudes of the coefficients are close to zero.

B3. Phone usage measures based on incoming and outgoing calls

The benchmark analysis is based on a directed network of outgoing calls. While the market share of our data provider is exceptionally high and varies little across space, the rational for limiting the anal-

Table B.2
Placebo Check: "Fake" Revision of Public Transport Schedules & Social Interactions, LPM-Models.

Sample Draw	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln(Travel Time $PT_{ij,t}$)	0.000	0.004	-0.010	-0.007	0.006	-0.009	-0.003	-0.011	-0.012	0.006
	(0.046)	(0.465)	(-1.162)	(-0.838)	(0.698)	(-0.965)	(-0.302)	(-1.230)	(-1.270)	(0.681)
Ln(Travel Time $PT_{ij,t}$)	0.119	0.076	-0.003	0.050	0.106	-0.155	0.069	-0.130	-0.122	0.062
	(0.862)	(0.459)	(-0.019)	(0.278)	(0.580)	(-0.976)	(0.439)	(-0.790)	(-0.947)	(0.443)
Ln(Travel Time $PT_{ij,t}$)^2	-0.014	-0.008	-0.001	-0.007	-0.011	0.017	-0.008	0.014	0.013	-0.006
	(-0.894)	(-0.450)	(-0.052)	(-0.329)	(-0.564)	(0.955)	(-0.474)	(0.748)	(0.895)	(-0.417)
Sample Draw	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Ln(Travel Time $PT_{ij,t}$)	-0.010	0.005	0.006	0.001	-0.004	-0.008	-0.006	-0.010	-0.007	-0.008
	(-1.079)	(0.559)	(0.690)	(0.076)	(-0.410)	(-0.803)	(-0.657)	(-1.024)	(-0.817)	(-0.823)
Ln(Travel Time $PT_{ij,t}$)	-0.105	-0.035	0.053	0.050	-0.067	0.015	-0.044	-0.111	-0.080	0.025
	(-0.724)	(-0.221)	(0.381)	(0.281)	(-0.363)	(0.097)	(-0.281)	(-0.728)	(-0.488)	(0.174)
Ln(Travel Time $PT_{ij,t}$)^2	0.011	0.005	-0.005	-0.006	0.007	-0.003	0.004	0.351	0.008	-0.004
	(0.685)	(0.263)	(-0.352)	(-0.287)	(0.355)	(-0.157)	(0.250)	(1.021)	(0.460)	(-0.235)

Notes: We use data from three-months windows prior to the change in the public transport timetable, i.e. June 2015–August 2015, and data for September 2015–November 2015 as for the placebo 'post-period'. Note that the actual change in the public transport timetable was implemented on December 13th, 2015 such that the specifications above test for trends in social interactions prior to the change i.e. whether differences could be observed already between August and September 2015. We estimate analogous specifications as in as in columns (5) and (6) of Table 5 and exclude all observations outside the canton of Zurich. We run these specifications for 20 samples following the sampling strategy described in Section 4.1. The samples covers only non-movers (both caller and callee) who used their phone every month at least once. All *coefficients* are multiplied by 10000, and can be interpreted as basis points. Standard errors are clustered by postcode pair and t-values are reported in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

Table B.3
Regional Differences in Phone Usage (*Incoming & Outgoing* Calls) of Movers, FE-Model.

a. Phone Usage Measures	Frequency		Duration		# Unique Contacts	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Pop. Density)	0.005		0.019***		-0.002	
	(0.004)		(0.005)		(0.003)	
City		0.034**		0.092***		0.001
		(0.011)		(0.016)		(0.008)
Adj. R^2	0.837	0.837	0.777	0.777	0.771	0.771
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Lang. Region & Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	90,011	90,011	90,011	90,011	90,011	90,011
b. Local Phone Usage Measures	Frequency		Duration		# Unique Contacts	
i.e. within 15min. Radius	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Pop. Density)	0.161***		0.367***		0.106***	
	(0.008)		(0.020)		(0.006)	
City		0.106***		0.150*		0.046***
		(0.025)		(0.062)		(0.017)
Adj. R^2	0.704	0.703	0.627	0.629	0.644	0.644
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Further Controls	Yes	Yes	Yes	Yes	Yes	Yes
Lang. Region & Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	90,011	90,011	90,011	90,011	90,011	90,011

Notes: We use monthly *incoming and outgoing calls* for June 2015–May 2016 to compute the logarithmized dependent variables. Note that population density is measured as the population within 15 minutes *road* travel time. The *sample* consists of movers who used their phone every month at least once. *Further controls* include commuting distance, dummy for belonging to language minority. Robust standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

ysis to outgoing calls was to ensure equal coverage of calls independent of the provider's local market penetration. Yet, the question remains whether the effects of population density on phone activity differ in terms of call direction. Accordingly, we reestimate our benchmark model using phone usage measures that combine outgoing and incoming calls.

The results are shown in Table B.3 and largely confirm the previous results. We obtain significant and positive effects of the city indicator on calling frequency and duration where the magnitude is similar to the corresponding effects reported in the benchmark analysis. Again, conditional on sorting there is no evidence of an effect on the number of unique contacts. With regard to the local measures, panel (b) of Table B.3 shows positive and significant effects for all three outcomes. The magnitude is well in line with the corresponding estimates for outgoing calls. Apparently, the effects of population density on phone usage are symmetric across outgoing and incoming calls.

B4. Heterogeneity

Table B.4
Heterogeneity in the Gravity Model of Social Interactions via Phone, LPM-Models.

Outgoing Calls	Pooled OLS: (1)	$\frac{\hat{\eta}_1}{Pr(g_{ij}=1\|T_{ij}=0)}$ (2)	OLS with Pair-FE: (3)	$\frac{\hat{\eta}_1}{Pr(g_{ij}=1\|T_{ij}=0)}$ (4)
Movers experiencing density increase				
Ln(Travel Time$_{ij,t}$)	-0.222	-0.454	-0.174	-0.389
Ln(Travel Time$_{ij,t}$)^2		0.050		0.042
Movers experiencing density decrease				
Ln(Travel Time$_{ij,t}$)	-0.207	-0.453	-0.176	-0.442
Ln(Travel Time$_{ij,t}$)^2		0.050		0.051
Movers of age 15–24				
Ln(Travel Time$_{ij,t}$)	-0.206	-0.453	-0.177	-0.400^a
Ln(Travel Time$_{ij,t}$)^2		0.050		-0.043^a
Movers of age 25–44				
Ln(Travel Time$_{ij,t}$)	-0.205	-0.454	-0.174	-0.427
Ln(Travel Time$_{ij,t}$)^2		0.050		0.048
Movers of age 45–64				
Ln(Travel Time$_{ij,t}$)	-0.204	-0.449	-0.174	-0.412
Ln(Travel Time$_{ij,t}$)^2		0.049		0.043
Female movers				
Ln(Travel Time$_{ij,t}$)	-0.205	-0.452	-0.176	-0.431
Ln(Travel Time$_{ij,t}$)^2		0.006		0.047
Male movers				
Ln(Travel Time$_{ij,t}$)	-0.204	-0.455	-0.171	-0.407
Ln(Travel Time$_{ij,t}$)^2		0.050		0.046

Notes: OLS-specifications as in columns (3) & (5) of Table 4, FE-specifications as in columns (7) & (9) of Table 4. To ease comparability we compute the distance elasticities adjusted for base probabilities ($\frac{\hat{\eta}_1}{Pr(g_{ij}=1|T_{ij}=0)}$) for each model and sample. All underlying estimates are at least significant on the 10 percent level except those indicated by a superscript a.

Table B.5
Heterogeneity Analysis for Regional Differences in Phone Usage Statistics, FE-Models.

***Local* Phone Usage Measures, i.e. within 15min. Radius**	# of Local Calls (1)	Duration of Local Calls (2)	# Unique Local Contacts (3)
Binary City Indicator			
Moves *into* city	-0.013	-0.125	0.029
	(0.032)	(0.092)	(0.023)
Moves *out of* city	0.153***	0.323**	0.126***
	(0.020)	(0.030)	(0.017)
Ln(Pop. Density)			
Movers experiencing density *increase*	0.094***	0.257***	0.072***
	(0.016)	(0.045)	(0.011)
Movers experiencing density *decrease*	0.169***	0.444***	0.117***
	(0.016)	(0.045)	(0.011)
Age Groups			
Movers of *age 15–24*	0.022^{+}	0.023	0.026**
	(0.013)	(0.036)	(0.009)
Movers of *age 25–44*	0.148***	0.356***	0.104***
	(0.010)	(0.028)	(0.007)
Movers of *age 45–64*	0.023***	0.684***	0.187***
	(0.018)	(0.053)	(0.013)
Gender			
Female movers	0.062***	0.189***	0.057***
	(0.010)	(0.027)	(0.007)
Male movers	0.176***	0.409***	0.126***
	(0.011)	(0.29)	(0.008)

Notes: FE-specifications as in columns (2), (5) & (8) of panel (b) in Table 7. $+p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

CRediT authorship contribution statement

Konstantin Büchel: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Maximilian v. Ehrlich:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

References

Arzaghi, M., Henderson, J.V., 2008. Networking off madison avenue. Rev. Econ. Stud. 75, 1011–1038.
Bailey, M., Cao, R., Kuchler, T., Stroebel, J., Wong, A., 2018. Social connectedness: measurement, determinants, and effects. J. Econ. Perspect. 32 (3), 259–280.
Barwick, P. J., Liu, Y., Patacchini, E., Wux, Q., 2019. Information, mobile communication, and referral effects. NBER Working Paper No. 25873.
Bailey, M., Farrell, P., Kuchler, T., Stroebel, J., 2020. Social conectedness in urban areas. J. Urban Econ. 118, 103264.
Berliant, M., Reed, R., Wang, P., 2006. Knowledge exchange, matching, and agglomeration. J. Urban Econ. 60, 69–95.
Blondel, V., Decuyper, A., Krings, G., 2015. A survey of results on mobile phone datasets and analysis. EPJ Data Sci. 4, 1–55.
Blumenstock, J., Eagle, N., Fafchamps, M., 2016. Airtime transfers and mobile communications: evidence in the aftermath of natural disasters. J. Dev. Econ. 120, 157–181.
Burley, J., 2015. The built environment and social interactions: Evidence from panel data. University of Toronto, mimeo.
Cairncross, F., 2001. The Death of Distance: How the Communication Revolution Is Changing Our Lives. Harvard Business School Press, Cambridge.
Catalini, C., 2017. Microgeography and the direction of inventive activity. Manage. Sci. 64 (9), 4348–4364.
Charlot, S., Duranton, G., 2006. Cities and workplace communication: some quantitative french evidence. Urban Stud. 43 (8), 1365–1394.
Ciccone, A., Hall, R., 1996. Productivity and the density of economic activity. Am. Econ. Rev. 86 (1), 54–70.
Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: sorting matters!. J. Urban Econ. 63, 723–742.
Combes, P.-P., Duranton, G., Gobillon, L., Roux, S., 2010. Estimating agglomeration effects with history, geology, and worker fixed effects. In: Glaeser, E. (Ed.), Agglomeration Economics. Chicago University Press, Chicago, pp. 15–65.
Kim, J., Patacchini, E., Picard, P., Zenou, Y., 2017. Urban Interactions. CEPR Discussion Paper No. 12432.
Catalini, C., Fons-Rosen, C., Gaulé, P., 2019. How do travel costs shape collboration? NBER Working Paper No. 24780.
ComCom, 2015. Tätigkeitsbericht der comcom 2015. Published online http://www.comcom.admin.ch/dokumentation/00564/index.html?lang=de (01.06.2016).
Cosslett, S., 1981. Maximum likelihood estimator for choice-based samples. Econometrica 49, 1289–1316.
Currarini, S., Jackson, M., Pin, P., 2009. An economic model of friendship: homophily, minorities, and segregation. Econometrica 77 (4), 1003–1045.
D'Costa, S., Overman, H., 2014. The urban wage growth premium: sorting or learning. Reg. Sci. Urban Econ. 48, 168–179.
Duranton, G., Puga, D., 2004. Micro-foundations of urban agglomeration economies. In: Henderson, J.V., Thisse, J. (Eds.), Handbook of Regional and Urban Economics. Elsevier, Amsterdam, pp. 2063–2115.
Eagle, N., Macy, M., Claxton, R., 2010. Network diversity and economic development. Science 328 (5981), 1029–1031. doi:10.1126/science.1186605.
Gaspar, J., Glaeser, E., 1998. Information technology and the future of cities. J. Urban Econ. 43, 136–156.
Graham, B., 2015. Methods of identification in social networks. Annu. Rev. Econ. 7, 465–485.
Helsley, R., Zenou, Y., 2014. Social networks and interactions in cities. J. Econ. Theory 150, 426–466.
Ioannides, Y., Overman, H., Rossi-Hansberg, E., Schmidheiny, K., 2008. The effect of information and communication technologies on urban structure. Econ. Policy 23 (54), 201–242.
Jackson, M., Rogers, B., 2007. Meeting strangers and friends of friends: how random are social networks? Am. Econ. Rev. 97 (3), 890–915.
Lancaster, T., 2000. The incidental parameter problem since 1948. J. Econom. 95 (2), 391–413.
Levy, M., Goldenberg, J., 2014. The gravitational law of social interaction. Physica A 393, 418–426.
Manski, C., Lerman, S., 1977. The estimation of choice probabilities from choice based samples. Econometrica 45 (8), 8.
Marmaros, D., Sacerdote, B., 2006. How do friendships form? Q. J. Econ. 121 (1), 79–119.
McPherson, M., Smith-Lovin, L., Cook, J., 2001. Birds of a feather: homophily in social networks. Annu. Rev. Sociol. 27, 415–444.
Milgram, S., 1967. The small-world problem. Psychol. Today 1 (1), 61–67.
Sato, Y., Zenou, Y., 2015. How urbanization affect employment and social interactions. Eur. Econ. Rev. 75, 131–155.
Schläpfer, M., Bettencourt, L., Grauwin, S., Raschke, M., Claxton, R., Smoreda, Z., West, G., Ratti, C., 2014. The scaling of human interactions with city size. J. R. Soc. Interface 11 (98), 1–9.
Shoag, D., Veuger, S., 2018. Shops and the city: evidence on local externalities and local government policy from big-box bankruptcies. Rev. Econ. Stat. 100 (3), 440–453.
Travers, J., Milgram, S., 1969. An experimental study of the small world problem. Sociometry 32 (4), 425–443.
Watts, D., 1999. Networks, dynamics, and the small-world phenomenon. Am. J. Sociol. 105 (2), 493–527.