Understanding the Role of Social Networks on Labor Market Outcomes Using a Large Dataset from a Mobile Network

Completed Research Paper[1]

Filipa Reis
Carnegie Mellon University
frodrigu@cmu.edu

Pedro Ferreira
Carnegie Mellon University
pedrof@cmu.edu

Abstract

We use a new and unique dataset combining social network data from Call Detail Records with employment information on mobile phone subscribers to study the role of information networks on job market outcomes. The novel contribution of our work is to focus on the effect of actual social connections beyond that associated to living in the same neighborhood. We find that the propensity to work together is two orders of magnitude greater for a pair of neighbors who call each other than that for a pair of neighbors who do not, suggesting that actual social ties play a significant role in learning about job opportunities. We also find that social networks play a stronger role in less privileged neighborhoods, which provides some evidence that social networks may be unable to mitigate the insulation problems of such neighborhoods.

Keywords: Job Information Networks; Call Detail Records

Introduction

In a world where Information and Communication Technologies (ICTs) are a pervasive aspect of daily life, who we know still defines, to a significant extent, what we know and when. The deliberate use of formal and informal social connections to access and exchange information is a strategy commonly used by anyone, and one that people tend to resort to when looking for a job (Ioannides and Loury 2004). Differences in how job related information is shared can play an important role in explaining differences in the labor market outcomes attained by people belonging to otherwise homogeneous socio-demographic groups (Cingano and Rosolia 2012).

One possible mechanism driving the relationship between social connections and labor market outcomes is the following: if employed people have inside information on employment opportunities, then the job seekers' ability to access this information will mostly depend on their connections to those employed. As a result, employed individuals will likely have a positive impact on the probability of their unemployed friends finding a job (Cingano and Rosolia 2012). On the other end, unemployed friends are unlikely to help finding a job. These asymmetric effects can potentially result in large discrepancies in long-term labor market outcomes arising from small temporaty employment shocks (Calvo-Armengol and Jackson 2004). These ideas have motivated scholars in sociology, economics and information systems to devote substantial effort to understand how social networks influence labor market outcomes (Ioannides and Loury 2004).

The general unavailability of actual social network data coupled with information on labor-market outcomes has limited the researchers' ability to empirically test the above mentioned mechanism. So far, most researchers have overcome this limitation using data from social experiments, such as re-location programs (Jacob 2003; Katz et al. 2000); self-reported data (Bentolila et al. 2010; Blau and Robins 1990), which in some cases are used to build social network proxies such as census block neighbors (e.g. Bayer et al. (2008)); and matched employer-employee data (Cingano and Rosolia 2012). Many of these data sources are only rarely produced and extremely costly to generate. More importantly, they do not contain any information on the actual social connections.

Today's widespread availability of large amounts of very detailed, and often real-time, data generated by ICTs creates opportunities to address the limitation above. In particular, the recent growth in the use of mobile communications and the subsequent availability of datasets on who calls whom, when and from where, opens up avenues for novel research approaches that may extend and complement the more traditional methodologies (Blondel et al. 2015). To the best of our knowledge, our paper is the first work that studies job information networks combining observational social network data (from mobile phone communications) with employment information (by using a sample of corporate paid mobile subscription plans).

We start by studying whether CDRs can be used to identify job information networks at the neighborhood level following the methodology developed by Bayer et al. (2008). We use a sample of mobile phone subscribers who have a corporate subscription plan paid by their employer. Combining the temporal call patterns and the geographical information on the cell-towers used to route calls we identify each subscriber's home and work location as the most used cell towers during nighttime and daytime, respectively. We then assess the propensity of two people living in the same location (cell-tower) to also work in the same location (cell tower), compared to the propensity of two people who live in the same neighborhood (parish) to do so. We interpret the difference between these two effects as an indication of the role that job information networks among people who live in close proximity may play in shaping job markets outcome. Our results show that, controlling for parish fixed effects, people who live in the same cell tower are 33 percent more likely to work in the same cell tower than pairs of subscribers who just live in the same parish. This result is in line with the findings of Bayer et al. (2008) and gives credibility to the use CDRs to study the role of job information networks at the neighborhood level.

Next we investigate the effect of actual social connections, beyond that associated to neighborhoods, on job market outcomes. This is operationalized by adding information on whether two individuals use their mobile phone to talk to each other to our model. We find that the propensity to work in the same location (cell tower) of two people who live in the same location (cell tower) and call each other is two orders of magnitude greater than that observed for people who live in the same cell tower but do not call each other. These results are robust to an alternative specification of the dependent variable – working for the same employer, which does not actually entail working in the same location – and also to different definitions of friendship. Here, our analysis adds a critical step to what has been previously done in the literature, which so far has been unable to account for the actual social connections among neighbors.

Finally, we look at whether the effect of social connections on job market outcomes varies with the socio-economic status of the place where people live. Previous research has suggested that job seekers from low income and low education neighborhoods tend to disproportionately rely on their neighbors, family, and friends to secure employment, a phenomena that may result in labor market insulation (Elliott 1999). Accordingly, we hypothesize that job information networks among neighbors will play a more important role in less privileged neighborhoods (Elliott 1999; Weinberg et al. 2004). Our results provide some evidence confirming this hypothesis and thus further the concerns regarding labor market insulation in economically worse-off neighborhoods, whose inhabitants are likely to have increased difficulty in assessing information on (better) job opportunities. This finding may call for policy measures targeted at improving access to labor market information in these areas.

The remainder of this paper is structured as follows. Section 2 reviews the extant literature on the role of social networks on labor market outcomes and on using CDRs for empirical research. Section 3 presents the empirical context for our study, our dataset, and our sampling strategy. Section 4 presents and discusses our analysis and results as well as the robustness checks we performed. Section 5 concludes.

Literature Review

Social Networks and Labor Market Outcomes

We can distinguish between two main sources of job related information – formal and informal (Rees 1966). Social networks are an informal source of job information that has received substantial attention by researchers in the fields of sociology, economics and information systems. The use of personal contacts in job search has been found to be prevalent and generally productive though outcomes vary according to individuals' socio-demographic characteristics (Holzer 1986; Ioannides and Loury 2004).

Social network features, such as size (Allen 2000), tie strength (Granovetter 1973), and diversity (Stoloff et al. 1999), have been found to play a significant role in providing people with new information on professional opportunities or even referrals, thus influencing i) participation in or withdrawal from the labor force (Calvó-Armengol and Jackson 2007); ii) employment status (Calvó-Armengol and Jackson 2007); iii) occupational choice (Bentolila, Michelacci, and Suarez, 2010), and iv) pecuniary and non-pecuniary compensations (Boxman et al. 1991; Franzen and Hangartner 2006). Both employed and unemployed individuals use social connections in job search, the latter enjoying a higher rate of job offers per contact than the former (Blau and Robins 1990).

Job information networks provide information and valuable insights not only to individuals but also to firms on potential candidates and job opportunities (Calvó-Armengol and Ioannides 2005), reducing both workers' and employers' uncertainty regarding the job's characteristics and the candidates' skills and competences (Montgomery 1991). Yet, some authors have argued that access to a network may not be necessarily associated to better labor market outcomes (Lin 1999; Stoloff et al. 1999). As employed individuals are likely to have inside information on employment opportunities, they can have a positive impact on the probability of their unemployed connections finding a job (Cingano and Rosolia 2012). Conversely, unemployed individuals will hold less information on job opportunities and thus may negatively impact the chances of their unemployed social connections finding a job. This mechanism can potentially result in large discrepancies in labor market outcomes of otherwise similar individuals arising from initially small employment shocks (Calvo-Armengol and Jackson 2004). Additionally, the effectiveness of social connections may vary widely across different economic actors, and while they are not necessarily the more effective job information channel, they tend to be predominantly used by the underprivileged and can make a significant difference in their outcomes (Lin 1999).

The general unavailability of actual social network data (which significantly limits researchers' ability to observe the features of the social networks under analysis (Stoloff et al. 1999)) for research and, in particular, social network data coupled with information on labor-market outcomes, has limited researchers' ability to empirically test the above mentioned mechanism. So far, most researchers have overcome this limitation using data from social experiments, such as re-location programs (Jacob 2003; Katz et al. 2000); self-reported data (Bentolila et al. 2010; Blau and Robins 1990), which in some cases are used to build social network proxies such as census block neighbors (e.g. Bayer et al. (2008)); or matched employer-employee data (Cingano and Rosolia 2012). However, many of these data sources are only rarely produced and extremely costly to generate and, more importantly, they do not contain any information on the actual social networks. Furthermore, the common use of survey data is likely to result in social network data that is biased towards the respondent's stronger or more frequent interactions (Campbell and Lee 1991).

Job Information Networks within Neighborhoods

Social connections among neighbors play a key role in one's life in fields as varied as health, education, labor or crime (Garner and Raudenbush 1991; Leventhal and Brooks-Gunn 2000; Sampson et al. 2002; Weinberg et al. 2004). The use and impact of these local networks has been found to differ according to the neighborhoods' social characteristics. In the case of the exchange of job related information, local social networks seem to be most relevant among less-educated workers and to have the strongest impact in less privileged neighborhoods (Weinberg et al. 2004). In particular, previous research has found that disadvantaged neighborhoods are often associated to labor market insulation (Elliott 1999). Namely, job seekers from low income and low education neighborhoods tend to disproportionately rely on their neighbors, family, and friends to

secure employment (Elliott 1999).

Research on neighborhood effects on labor market outcomes has developed empirical approaches for studying the effects of local social connections on socio-economic outcomes that do without information on the actual social connections. One of these approaches is the use of data from social experiments in which individuals and/or families relocated from poor neighborhoods to neighborhoods with higher standards of living (Katz et al. 2000; Ludwig 1999). Yet, it is hard for researchers in this field of study to effectively define control and treatment groups. Namely, the treated individuals or families are typically those that pre-qualify for re-location and this selection problem hampers significantly the researchers' ability to identify the effect of treatment (Bayer et al. 2008).

Another approach, which deals with potential correlation across unobserved characteristics of the individuals living in the same neighborhood, is to use population survey data aggregated at higher geographical levels (Ross 1998). However, this methodology averages effects across neighborhoods in the same area, which may experience outcomes with very different magnitudes thus precluding researchers from understanding exactly which mechanisms perform better and where (e.g. high poverty and high income areas behave differently) (Bayer et al. 2008). A way to overcome these limitations is to disaggregate the data to the lowest possible geographical level so that the variation in the characteristics of the residents within the same neighborhood can be isolated (Bayer et al. 2008). This approach is followed by Bayer et al. (2008) who use detailed data from the Boston Census to show evidence of job referrals in neighborhoods. The authors find that pairs of workers who live in the same block are 33 percent more likely to work in the same block than pairs of workers who live in the same group of blocks but not exactly in the same block.

The Use of Call Detail Records in Research

The availability of big datasets generated by ICTs allows researchers from various fields to understand behavior from direct observation instead of elicitation through surveys, which, for the most part, suffer from self-report bias. While one cannot expect big datasets to reflect the full extent of social interactions or that of individuals' attitudes and behaviors, they nonetheless allow researchers to surpass many of the limitations associated to traditional survey-based methodologies (Hidalgo and Rodriguez-Sickert 2008). As such, they constitute a source of data that can be used as a valuable complement or alternative to many traditional research methods.

The ubiquitous use of mobile phones across all strata (individuals of all genders, ages, and social classes use either smart or dumb phones) makes Call Detailed Records (CDRs) one of the most comprehensive sources of data on an individual's social interactions (Hidalgo and Rodriguez-Sickert 2008) and patterns of spatial mobility (Calabrese et al. 2010b; Song et al. 2010). CDRs have been shown to accurately predict friendships (Eagle et al. 2009), and phone calls have been shown to support and reinforce the same social network as that established through face-to-face communications (Kim et al. 2007). In a recent work, Blondel et al. (2015) survey the results on the study of mobile phone datasets; the developments in this field have mostly concerned social network analysis and social network analysis across time and space (dynamical networks and geographical networks) and its main practical applications concern urban sensing, epidemics, health monitoring, viral marketing, crime detection, and development (Blondel et al. 2015).

A common pre-requisite to address interesting socio-economic questions is the availability of data on the demographic characteristics of individuals. This is a key limitation of CDRs, which generally lack demographic information about consumers due to non-collection and/or privacy reasons. Several studies using CDRs have partially overcome this limitation by inferring the residential and/or workplace locations of mobile phone subscribers through the combination of the temporal call patterns and the geographical information on the cell towers used to route the calls (Baker and Fradkin 2014; Calabrese et al. 2010a; Cici et al. 2013; Isaacman et al. 2011; Onnela et al. 2011; Phithakkitnukoon et al. 2011a; Phithakkitnukoon et al. 2011b; Verkasalo 2009). A common approach is to classify the most used cell towers during nighttime and daytime as home and work, respectively (e.g. Kirkpatrick et al. (2012) and Phithakkitnukoon et al. (2011b)) and then validate these estimations by ascribing each subscriber to the municipality where their home cell tower is located and then correlating the number of subscribers per municipality with national census data (Phithakkitnukoon et al. 2011b; Tizzoni et al. 2014). In this work we will follow a similar approach to identify

and validate home and work locations of mobile phone subscribers.

To date, there is still limited research using CDRs for socio-economic analysis. This is likely due to the lack of demographic information and limited access of ground-truth data for testing the quality of predictions. A couple of exceptions are the use of CDRs to investigate the relationship between the structure of social networks in geographical regions and the region's level of economic development (Eagle et al. 2010), to predict the level of economic development in an urban setting (Soto et al. 2011), and to predict employment shocks (Toole et al. 2015). To our knowledge, Toole et al. (2015) work is so far the only study using CDRs for studying labor market outcomes. These authors find that the use of micro-level mobile call features aggregated at the regional level improves upon the existing unemployement rate forecasts (Toole et al. 2015).

Hypotheses

Given the limitations inherent to Call Detail Records highlighted in the previous section, a fundamental first step in our analysis is to verify whether CDRs allow us to identify an effect similar to that observed in Bayer et al. (2008) before we attempt to tease out the effect of the actual social connections from the effect of being neighbors. This brings us to our first hypothesis:

Hypothesis 1: People who live in the same cell tower are more likely to work in the same cell tower than people who live in the same parish but not in the same cell tower.

Support for hypothesis 1 can be interpreted as evidence of information exchange on job opportunities and job referrals among neighbors resulting from the interpersonal contact caused by living close by. However, not all people who live in the same region will know or talk to each other and the effect measured under this hypothesis will necessarily average out the effect of people who actually know each other and communicate and people who do not. Information sharing is most likely to occur among people who actually know each other (though we do not exclude the possibility that it can occur indirectly through second degree connections). Therefore, we posit that actual social connections mediate the relationship between being neighbors and being co-workers.

Hypothesis 2: People who live in the same cell tower and call each other are more likely to work in the same cell tower than people who live in the same cell tower but who do not call each other.

Finally, we hypothesize that the socio-economic conditions of a given area will moderate the impact that social ties among neighbors may have on the likelihood of working together. People in more privileged neighborhoods are likely to have better access to information and, in particular, more access to more (and perhaps better) sources of information on job opportunities. Social networks are thus likely to play a more important role in sharing information about job opportunities in less privileged neighborhoods as has been suggested by previous research (Elliott 1999). Thus, we expect the impact of social connections on the likelihood of working in the same location to be greater for pairs of neighbors who reside in areas of lower socio-economic status relative to pairs of neighbors residing in areas that are better off, which brings us to our third hypothesis.

Hypothesis 3: The propensity of two people living in the same cell tower and who call each other to work in the same cell tower is higher for those who live in economically worse off parishes relative to those who live in economically better off parishes.

Empirical Context, Data and Sampling Strategy

Our dataset contains the Call Detail Records of all subscribers of a major mobile phone carrier in the largest metropolitan region in one European country, between August 2008 and June 2009. During this period, 90 percent of the country's population had a mobile phone and three main carriers dominated the market. In our study, we use CDRs from the market leader in the corporate segment.

The region under analysis is nearly 3,000 square kilometers and houses approximately 3 million inhabitants. This region is divided into 18 municipalities, which are in turn subdivided into 211 civil parishes. Civil parishes are the lowest geographical level at which census data is available and were originally established as ecclesiastical divisions – inhabitants of the same parish used to attend the same church.

There are 4,909 cell-towers in the metropolitan region under study. Thus, on average, each parish has 23 cell-towers. The coverage area of each cell tower varies according to the density of cell towers in the region and to the characteristics of the landscape. In urban regions it may span less than 1 square kilometer while in rural areas it may exceed 3 square kilometers (Soto et al. 2011). We used an API provided by Google to find the GPS coordinates and the parishes where cell towers are located.

Our data includes roughly 4.5 million users and 3.7 billion calls. Each CDR identifies the caller and callee, the cell towers used to route the call and respective geographical coordinates, the time at which the call was placed and its duration. Most subscribers in our data are assigned to a customer segment. For the purpose of our analysis, we group customer segments into two main groups: i) business – all subscribers with an employer (company) paid plan, most of which have a post-paid tariff plan; ii) non-business – all other segments, the vast majority of which have a pre-paid tariff plan. Note that we can identify business subscribers working for the same company as those who belong to the same corporate account. In total, there are 336,198 business subscribers who were active during the 11-month period of analysis.

Using the temporal call patterns of phone calls, we identified the home and work locations of each subscriber as the most used cell tower between 7pm-7am every day of the week and 1pm-5pm on weekdays, respectively (time windows based on textciteisaacman2011identifying), during the entire period of analysis. We excluded from our dataset all subscribers for whom we were not able to uniquely identify one home and one work location, distinct from each other, and both located in the metropolitan area considered. Note that subscribers that live and work in the same place might do so not because they influence each other to work in the same location but because of reasons unobserved to us, thus we do not consider them in our study. We also remove from our sample subscribers associated to corporate accounts with less than five people, which may represent families, rather than corporations, who obtain a corporate account to get a better deal (unlike other mobile carriers, this carrier did not offer family plans).

Limiting the geographical scope of our analysis to a single metropolitan area allows us to have a sufficiently high number of individuals per cell tower even when using a relatively small sample of subscribers. However, we should note that our sampling process eliminates from our sample: 1) all workers who work either at home or very close by (this is not different from the approach taken by Bayer et al. (2008) in which people who worked in the same block where they lived were not considered); 2) all workers who do not have a business subscription. Therefore, we point out, as a limitation, that our findings may not generalize beyond the specific type of worker considered in our study.

After our sample selection criteria are implemented, we are left with a subset of 20,748 corporate subscribers that satisfy all of the above criteria. For computational reasons, we randomly selected a smaller subset of 15,000 subscribers for our analysis. T-tests show that the subscribers in our sample are significantly more active in terms of cell phone usage than the average corporate subscriber as depicted in the boxplots in Figures 1 to 3. This is likely due to the fact that these subscribers live and work in the largest metropolitan region in the country, where the largest business center is also located, which may bias our sample towards the more active corporate users.

The subscribers in our final sample live in 2,720 cells towers, located in 196 parishes out of the 211 in this metropolitan area, and work in 2,266 cells towers, located in 195 parishes out of the 211 in this metropolitan area. Each cell tower where people live has on average 6 subscribers and each cell tower where people work has on average 7 workers. Figure 4, shows that the number of subscribers per parish in our sample is aligned with the population per parish and therefore our sampling strategy does not seem to introduce significant bias in this respect.

Next, we generate all possible pairs of subscribers using our sample of 15,000 subscribers. For each pair of subscribers, a dummy variable is created indicating whether the two members live in the same parish. Only pairs of subscribers who live in the same parish are considered in our analysis. This leaves us with a total of 1,495,157 pairs of subscribers. Then, for each pair of subscribers, a dummy variable is created indicating whether the subscribers in the pair live in an area covered by the same cell tower – that is, whether they are neighbors –, and two other indicating whether they work in the area covered by the same cell tower or for the same employer – or whether they are co-workers.
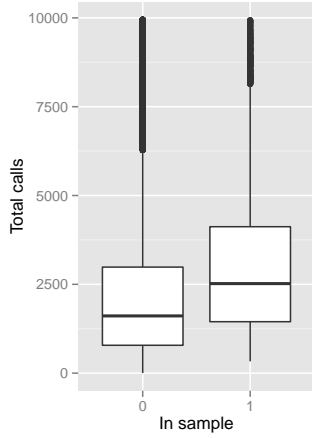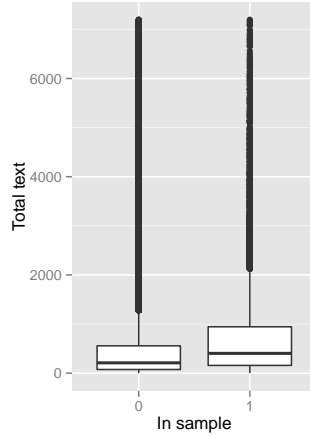
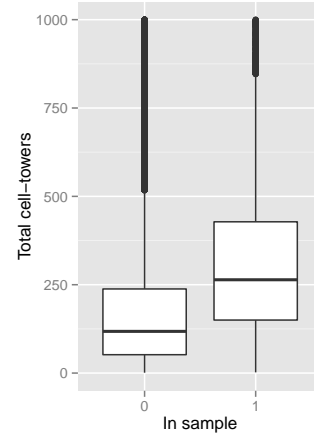**Figure 1. Calls**   **Figure 2. Text**   **Figure 3. Cells**
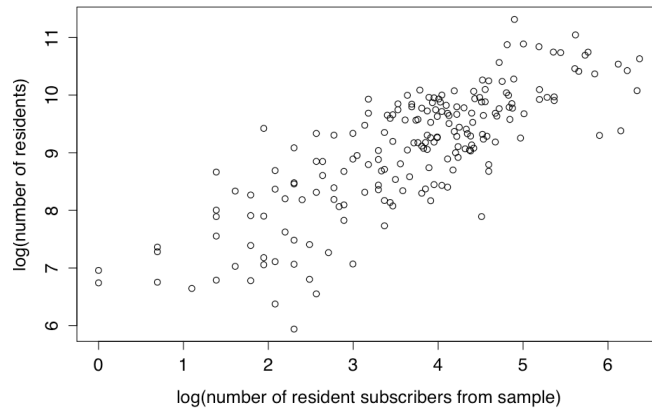

Figure 4. Relationship between true parish population and sample's parish population

We note that our classification of pairs of subscribers as neighbors (or location based co-workers) contains measurement error. For example, two subscribers that live in distant opposing ends of the area covered by a cell tower may be incorrectly classified as neighbors and two subscribers who live close to each other but in the coverage areas of two distinct cell towers may each be directed to a different cell tower and hence incorrectly classified as non-neighbors. There is, however, no reason to expect this error to be biased in one direction or another and, therefore, we expect its impact on our results to only attenuate the actual effect that living in the same location may have on the propensity of two subscribers to work in the same location.

Finally, for each subscriber pair, we create an indicator variable of whether or not the two members of the pair are friends. One can employ a number of different strategies to infer friendships from CDR data (see, for example, Godinho de Matos et al. (2014)). In our analysis we consider a strict definition of friendship requiring that the number of calls exchanged between the members of the pair must be above the median number of calls exchanged by the subscriber pairs in our sample with a positive number of calls exchanged. Alternative definitions of friendship are used in our robustness checks and the corresponding results are presented in the appendix. Table 1 presents some pair level descriptive statistics.

		Live in the same Parish but not in the same Cell		Live in the same Parish and in the same Cell	
		N = 1,135,530		N = 359,627	
		Work in the same cell	Work for the same company	Work in the same cell	Work for the same company
	N	6873	4484	2114	1929
Calls >0	3469	545	1436	388	818
Calls reciprocal >0	2529	416	1111	336	706
Text messages >0	2288	364	979	307	658
Calls >0 & Text messages >0	2174	347	940	302	648
Calls 7pm-7am >0	2111	317	877	295	630
Calls >= median	1137	284	774	281	603

Number of pairs = 1,495,157

Table 1. Pair level descriptive statistics

Table 2 shows the correlation matrix for the main covariates used in our study. The correlations show that working in the same company is more correlated with being friends than just working in the same cell tower. Also, the variables capturing the different definitions of friendship are highly correlated. Table 3 shows descriptive statistics of the mobile phone activity of the users in our final sample.

		1	2	3	4	5	6	7	8	9
1	**same work cell**	1.000								
2	**same employer**	0.274	1.000							
		0.000								
3	**same home cell**	-0.001	0.009	1.000						
		0.238	0.000							
4	**I(calls >0)**	0.164	0.476	0.010	1.000					
		0.000	0.000	0.000						
5	**I(calls reciprocal >0)**	0.155	0.450	0.011	0.854	1.000				
		0.000	0.000	0.000	0.000					
6	**I(text messages >0)**	0.146	0.426	0.012	0.771	0.817	1.000			
		0.000	0.000	0.000	0.000	0.000				
7	**I(calls >0 & text messages >0)**	0.144	0.424	0.013	0.791	0.839	0.975	1.000		
		0.000	0.000	0.000	0.000	0.000	0.000			
8	**I(calls 7pm-7am >0)**	0.138	0.408	0.013	0.780	0.819	0.785	0.805	1.000	
		0.000	0.000	0.000	0.000	0.000	0.000	0.000		
9	**I(calls >median)**	0.141	0.411	0.014	0.708	0.821	0.806	0.827	0.810	1.000
		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

Table 2. Correlation table

Variable	Obs.	Mean	Std. Dev.	Minimum	Maximum
Distance between home and work (km)	14,998	10.37	8.53	0	62
Total number of calls	14,998	3,289.25	2,617.34	332	29,802
Total number of text messages	14,998	925.42	1,616.17	2	52,095
Number of ties	14,998	248.99	222.09	3	7,713
Number of workers in company	14,998	122.74	224.18	5	1,605

Table 3. User level descriptive statistics

Empirical Strategy

We start by testing our first hypothesis and analyzing the propensity of two neighbors in our sample to work in the same location (cell tower coverage area), analogously to Bayer et al. (2008). Our first empirical specifications are:

$$W_{ij}^{cell} = \beta_0 + \beta_1 N_{ij}^{cell} + \epsilon_{ij} \tag{1}$$

$$W_{ij}^{cell} = \rho_{parish} + \beta_1 N_{ij}^{cell} + \epsilon_{ij} \tag{2}$$

where W_{ij} indicates whether subscribers i and j work in the same cell tower and N_{ij} indicates whether they are neighbors. In (1) β_0 captures the baseline propensity for two subscribers who live in the same parish (but are not neighbors) to work in the same location and β_1 shows how much higher this propensity is for neighbors (subscribers that not only live in the same parish but also in the same cell tower). In (2), we add parish-level fixed-effects, ρ, to control for potentially time-invariant unobserved parish characteristics, such as access to public transportation, which may lead to correlation in unobserved attributes across neighbors. In all regressions, we cluster errors at the parish level. In (2), β_1 estimates the effect of being neighbors on the propensity to work in the same location.

Next, in order to test hypothesis 2, we extend upon the model above by adding F_{ij}, an indicator of whether subscribers i and j are friends (using our stricter definition of friendship) and then, an interaction term of neighbors and friends.

$$W_{ij}^{cell} = \rho_{parish} + \beta_1 N_{ij}^{cell} + \beta_2 F_{ij} + \epsilon_{ij} \tag{3}$$

$$W_{ij}^{cell} = \rho_{parish} + \beta_1 N_{ij}^{cell} + \beta_2 F_{ij} + \beta_3 N_{ij}^{cell} F_{ij} + \epsilon_{ij} \tag{4}$$

In (3) β_2 captures the effect of being friends controlling for being (or not) neighbors. Note that the correlation between being friends and living in the same location is only 0.014. We are particularly interested in the output of model (3) to learn whether being friends impacts the propensity to work in the same location beyond the effect of being neighbors. In (4) β_3 captures how the effect of being friends on the propensity to work in the same location changes for neighbors versus people who are not neighbors. Friends may influence each other about job opportunities irrespective of where they live. A positive β_3 will indicate that being friends increases the propensity to work in the same location across neighbors more than across pairs of people that live further apart. This may happen if, for example, communication across friends reinforces the information about job opportunities that people already share because they are neighbors.

Finally, we test our third hypothesis that importance of social connections will be stronger among people who live in disadvantaged neighborhoods. For this part of our analysis, we add to our model data on the socioeconomic level of each parish, namely the average cost of mortgage and the unemployment rate from the region's census. The former is a good measure for the socioeconomic conditions of a region (Singh, 2003). It proxies income well and is available at the parish level for the metropolitan region we study in this

paper. [2] We use CM_{parish} and UR_{parish} to represent the cost of mortgage and the unemployment rate of a parish, respectively. We interact these covariates with the effect of friendship under two new specifications:

$$W_{ij}^{cell} = \rho_{parish} + \beta_1 N_{ij}^{cell} + \beta_2 F_{ij} + \beta_3 N_{ij}^{cell} F_{ij} + \beta_4 CM_{parish} F_{ij} + \epsilon_{ij} \tag{5}$$

$$W_{ij}^{cell} = \rho_{parish} + \beta_1 N_{ij}^{cell} + \beta_2 F_{ij} + \beta_3 N_{ij}^{cell} F_{ij} + \beta_4 UR_{parish} F_{ij} + \epsilon_{ij} \tag{6}$$

Note that CM_{parish} and UR_{parish} are parish-level effects so their direct effects are captured in the parish dummies. In both (5) and (6) a statistically significant β_4 indicates that the effect of friendship on working in the same location is different across socioeconomic strata.

All models are estimated using ordinary least square. The main reasons to use a linear probability model (LPM) in this paper follow the arguments set forth in Angrist and Krueger (2001) and Wooldridge (2002), ch. 15: i) it is consistent and readily yields marginal effects; ii) it is appropriate when independent variables are binary (Wooldridge (2002), pp. 475). Although LPM estimates should be seen as convenient approximations for the underlying response probability, if one's goal is to estimate a partial effect, either at the average of the independent variable, or their average across the distribution of the independent variable, the fact that some predicted values may be off is typically not important. Heteroskedasticity will always be present in LPM but we address this issue by using a heteroskedasticity robust calculation of standard errors.

Results, Robustness Checks and Discussion

Main Results

Table 4 shows results for the first four models specified in the previous section. Column (1) shows us that the propensity for a pair of subscribers who live in the same parish but not in the same cell tower to work in the same location is 0.61 percent. This estimate has the same order of magnitude of that found by Bayer et al. (2008). Column (2) shows us that the propensity of two subjects that live in the same location to work in the same location is 0.20 percent. This result is obtained after controlling for parish fixed-effects and is also of the same order of magnitude of that obtained by Bayer et al. (2008) - living in close proximity increases the probability of working together for two people living in the same parish by 33 percent . Column (3) shows that being friends plays an important role in the propensity to work in the same location – it increases by 31.52 percent when people are friends. This result suggests that social connections play a much more important role in determining labor market outcomes than the neighborhood one lives in (although this continuous to play a significant role even when social connections are accounted for.) Column (4) shows evidence that the propensity to work in the same location is even higher for neighbors who are also friends, namely 0.0011+27.09+10.63=37.83 percent .

These results provide supporting evidence for both hypothesis 1 and hypothesis 2. Although geographic proximity still has a very strong effect on the propensity to work together, friendship seems to increase this propensity further, leading us to believe that friends exchange information about job opportunities that ultimately may lead them to end up working in the same location. This result should, however, be interpreted with caution because alternative explanations, such as homophily and reverse causation, may be driving our results. The robustness checks presented in the next section attempt to alleviate these concerns.

Table 5 shows the results for the last two models specified in the previous section. Columns (1) and (2) pertain to working in the same location whereas Columns (3) and (4) pertain to working in the same company. As expected, β_4 is negative in the case of mortgage costs and positive in the case of unemployment rate (always statistically significant except for the case of unemployment rate when using working in the same company as dependent variable). This confirms the hypothesis that social connections among employed people (recall that our analysis considers only workers) have a stronger impact in neighborhoods with lower

---

[2]The region considered, being the largest metropolitan region in the country, exhibits a cost of mortgage significantly higher than the national average. As for unemployment, the average unemployment rate in the metropolitan area that we analyzed is similar to the national average, though it varies less.

	(1) work cell	(2) work cell	(3) work cell	(4) work cell
neighbors	-0.0002 (0.0001)	0.0020*** (0.0002)	0.0013*** (0.0002)	0.0011*** (0.0002)
friends			0.3152*** (0.0112)	0.2709*** (0.0141)
neighbors and friends				0.1063*** (0.0229)
intercept	0.0061*** (0.0001)	- -	- -	- -
Parish	no	yes	yes	yes
Observations	1495157	1495157	1495157	1495157
R^2	0.000	0.007	0.026	0.027

Standard errors in parentheses
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 4. Regression results

income and with higher unemployment rates. This is expected because access to formal employment channels may be less pervasive in these neighborhoods.

The above also reinforces the underlying idea that the information shared about job opportunities in underprivileged neighborhoods might not necessarily lead to significant improvements in employment conditions. Thus, these findings provide some evidence in support of policies such as providing housing opportunities for poorer families in higher income neighborhoods. As argued in previous literature, for more disadvantaged individuals, social mobility may require one to secure access to "resources beyond the usual circles" (Lin 1999)). Relocation programs are one possibly effective mechanism to help them establish new social connections that, as we show in this paper, may eventually bring them in touch with opportunities to improve their jobs and consequently their social status.

	(1) work cell	(2) work cell
neighbors	0.0011*** (0.0002)	0.0011*** (0.0002)
friends	0.5169*** (0.0743)	0.0498 (0.0608)
neighbors and friends	0.1000*** (0.0229)	0.1044*** (0.0228)
mortage x friends	-0.0007*** (0.0002)	
unemp. rate x friends		0.0184*** (0.0050)
intercept	- -	- -
Parish	yes	yes
Observations	1495157	1495157
R^2	0.027	0.027

Standard errors in parentheses
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 5. Regression results - parish socio-economic level

Robustness checks

Our main empirical specifications allow us to correlate being neighbors and being friends with the propensity to work in the same location. However, our ability to claim a causal direction in our results is limited by the possibility of reverse causality, namely, two people may live in the same place as a result of working in the same place and not the other way around.

Distance between home and work

On the one hand, coworkers may share information on housing opportunities among them and later become neighbors as a consequence of that information exchange (be it close to their work site or far from it); on the other hand there may be corporate housing, or housing advertisements at the company location, or any other kind of on-site housing promotion that provides workers with information about housing opportunities that are conveniently located close to the company's site. As a first robustness check, we use the distance between each subscribers home and work location to address the second part of the problem as we consider it most likely that housing opportunities promoted at the company site are mostly or exclusively of houses located in great proximity to the company's location. Though this does not fully address the reverse causality problem we hope that it significantly reduces it.

We operationalize this robustness check by interacting a dummy variable called "both far", indicating whether both subscribers in a pair live more than 10 Km from where they work, with the variables of interest. The threshold was chosen based on the mean of the distribution of distance between home and work; the median being 7.8 Km. This dummy variable is 1 for 282,142 pairs of subscribers in our sample. The results, summarized in Table 6 in the appendix, show that the effect of friendship on the propensity to work in the same location does not vanish for subscribers who work far from where they live. We also ran our main regressions for only these subset of 282,142 pairs of subscribers and the results remain as before and are available upon request.

Definition of friends

In their study on job search on online social networks, Garg and Telang (2011) find that strong friends are more likely to help an individual get a job while weak friends are more likely to provide an individual with new information about job opportunities but not the actual job. To address the possibility that different definitions of friendship may yield different results we repeated our analysis using multiple of alternative definitions of friendship. These alternate definitions include two-way communication - pairs of subscribers who have exchanged reciprocal calls –, and night-time and weekend call exchanges – aims at reducing our concerns with reverse causality as we expect people who call each other outside work hours to be more likely to be actual friends. These results for different definitions of friendship are summarized in Tables 7 and 8 in the appendix and are consistent with the main results presented in the paper.

Definition of co-workers

Influence across friends may result in working in the same company besides working in the same location. Therefore, in another robustness check, we will replace our dependent variable by whether subscribers i and j work in the same company, C_{ij}, which, in fact, may not imply working in the same location[3]. Subjects might be better informed about job opportunities in the company they work for and thus might be more likely to spread this information better and faster. To accomplish this test we change the definition of the dependent variable to indicate whether subscribers i and j work for the same company. The results are summarized in Table 8 in the appendix and are consistent with the previous findings.

Other checks

Other robustness checks conducted include adding controls on the call activity of each member of the pair to our models - number of calls, number of text messages, number of cells contacted, and number of ties; and

---

[3]In our sample of 15,000 business subscribers, the location with the highest number of workers has on average 70.52 percent of the workers associated to that company.

re-running our analysis with a user sample from which we only exclude subscribers who work for companies with only one worker. The results from both analyses remain qualitatively unchanged and are available upon request.

Conclusion

We use eleven months of Call Detail Records (CDRs) from one major mobile carrier in a large European metropolitan area to empirically detect job information networks at the neighborhood level extending Bayer et. al.'s (2008) empirical approach by adding social interactions across subscribers via mobile phone. Our analysis starts by showing that subscribers who live in the same location are more likely to work in the same location. We then add to this formulation information about who calls whom. Our results recover previous findings, namely, that living in the same location has a positive and significant effect on the propensity to work in the same location. We find that friendships – proxied by mobile calls – across neighbors also play a significant role in the propensity to work together in the same location. The propensity of neighboring friends to work in the same location is two orders of magnitude above that associated with being only neighbors. This finding provides some evidence that local information networks about job opportunities play a significant role in the processes of job search and hiring.

We provide a number of robustness checks aimed at alleviating concerns with homophily and reserve causality. In particular, we show that our results remain unchanged when we classify as friends subscribers who call each other at night, thus outside the work schedule, which is more likely to only pick up personal friendships, or subscribers that call each other 2-ways (each of them initiates a call), which is likely to avoid spurious connections such as those one maintains with voicemails or call centers. In another robustness check, we show that our results also remain unchanged for subscribers who live far away from where they work. This allows us reduce our concerns regarding the fact that some subscribers may become friends and neighbors because they work for the same company that, for example, provides them subsidized housing, which is typically located close to the workplace. We also show that our results remain unchanged when we use working for the same company as a dependent variable in lieu of (just) working in the same location.

Finally, we present results interacting friendship with socioeconomic characteristics of the parishes studied, namely, mortgages costs and unemployment rates. We show that the effect of friendships on working in the same location, or in the same company, is stronger in neighborhoods with lower mortgage costs and higher unemployment rates. This finding provides some evidence that social networks may not help underprivileged individuals improve their job status because their friends and neighbors might be unable to share interesting job opportunities or provide valuable job referrals. Thus, re-allocation programs aimed at moving poorer families to more affluent neighborhoods may help limit this insulation effect.

While our work is among the first to use a large dataset on call detailed records for social economic analysis our data are still limiting. CDRs are a wealthy source of data to infer one's social network. However, they provide little demographic information. The latter can only be obtained by matching where subscribers spend most of their time (inferred from the cell towers they use) to census data, which results in obtaining data only at a higher level of aggregation, precluding use from exploring individual-level heterogeneity in socioeconomic covariates. Yet, our paper shows how CDR data, which are already and passively collected by mobile carriers, can be used to perform social economic research, thus without additional data collection costs, which are massive in the case of census data. Therefore, we hope that our paper opens up the way for future research that leverages existing mobile data sources to study human behavior and interactions across human subjects.

References

Allen, W. D. 2000. "Social networks and self-employment," *The Journal of socio-economics* (29:5), pp. 487–501.

Angrist, J. and Krueger, A. B. 2001. *Instrumental variables and the search for identification: From supply and demand to natural experiments*. Tech. rep. National Bureau of Economic Research.

Baker, S. R. and Fradkin, A. 2014. "The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data," *Available at SSRN 2251548*.

Bayer, P., Ross, S. L., and Topa, G. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy* (116:6), pp. 1150–1196.

Bentolila, S., Michelacci, C., and Suarez, J. 2010. "Social contacts and occupational choice," *Economica* (77:305), pp. 20–45.

Blau, D. M. and Robins, P. K. 1990. "Job search outcomes for the employed and unemployed," *Journal of Political Economy*, pp. 637–655.

Blondel, V. D., Decuyper, A., and Krings, G. 2015. "A survey of results on mobile phone datasets analysis," *arXiv preprint arXiv:1502.03406*.

Boxman, E. A., De Graaf, P. M., and Flap, H. D. 1991. "The impact of social and human capital on the income attainment of Dutch managers," *Social networks* (13:1), pp. 51–73.

Calabrese, F., Lorenzo, G. D., and Ratti, C. 2010a. "Human mobility prediction based on individual and collective geographical preferences," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on,* IEEE, pp. 312–317.

Calabrese, F., Pereira, F. C., Di Lorenzo, G., Liu, L., and Ratti, C. 2010b. "The geography of taste: analyzing cell-phone mobility and social events," in *Pervasive computing,* Springer, pp. 22–37.

Calvó-Armengol, A. and Ioannides, Y. M. 2005. *Social networks in labor markets.* Tech. rep. Department of Economics, Tufts University.

Calvo-Armengol, A. and Jackson, M. O. 2004. "The effects of social networks on employment and inequality," *American economic review*, pp. 426–454.

Calvó-Armengol, A. and Jackson, M. O. 2007. "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of economic theory* (132:1), pp. 27–46.

Campbell, K. E. and Lee, B. A. 1991. "Name generators in surveys of personal networks," *Social networks* (13:3), pp. 203–221.

Cici, B., Markopoulou, A., Frías-Martínez, E., and Laoutaris, N. 2013. "Quantifying the potential of ridesharing using call description records," in *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications,* ACM, p. 17.

Cingano, F. and Rosolia, A. 2012. "People I know: job search and social networks," *Journal of Labor Economics* (30:2), pp. 291–332.

Eagle, N., Macy, M., and Claxton, R. 2010. "Network diversity and economic development," *Science* (328:5981), pp. 1029–1031.

Eagle, N., Pentland, A. S., and Lazer, D. 2009. "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences* (106:36), pp. 15274–15278.

Elliott, J. R. 1999. "Social isolation and labor market insulation," *The Sociological Quarterly* (40:2), pp. 199–216.

Franzen, A. and Hangartner, D. 2006. "Social networks and labour market outcomes: The non-monetary benefits of social capital," *European Sociological Review* (22:4), pp. 353–368.

Garg, R. and Telang, R. 2011. "To Be or Not To Be Linked on LinkedIn: Job Search Using Online Social Networks," *Available at SSRN 1813532*.

Garner, C. L. and Raudenbush, S. W. 1991. "Neighborhood effects on educational attainment: A multilevel analysis," *Sociology of education*, pp. 251–262.

Godinho de Matos, M., Ferreira, P., and Krackhardt, D. 2014. "Peer influence in the diffusion of the iPhone 3G over a large social network," *Management Information Systems Quarterly (Forthcoming)*.

Granovetter, M. S. 1973. "The strength of weak ties," *American journal of sociology*, pp. 1360–1380.

Hidalgo, C. A. and Rodriguez-Sickert, C. 2008. "The dynamics of a mobile phone network," *Physica A: Statistical Mechanics and its Applications* (387:12), pp. 3017–3024.

Holzer, H. J. 1986. *Search method use by unemployed youth.* Tech. rep. National Bureau of Economic Research.

Ioannides, Y. M. and Loury, L. D. 2004. "Job information networks, neighborhood effects, and inequality," *Journal of economic literature*, pp. 1056–1093.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., and Varshavsky, A. 2011. "Identifying important places in people?s lives from cellular network data," in *Pervasive computing,* Springer, pp. 133–151.

Jacob, B. A. 2003. *Public housing, housing vouchers and student achievement: Evidence from public housing demolitions in Chicago.* Tech. rep. National Bureau of Economic Research.

Katz, L. F., Kling, J. R., and Liebman, J. B. 2000. *Moving to opportunity in Boston: Early results of a randomized mobility experiment.* Tech. rep. National Bureau of Economic Research.

Kim, H., Kim, G. J., Park, H. W., and Rice, R. E. 2007. "Configurations of relationships in different media: FtF, email, instant messenger, mobile phone, and SMS," *Journal of Computer-Mediated Communication* (12:4), pp. 1183–1207.

Kirkpatrick, S., Kulakovsky, A., Cebrian, M., and Pentland, A. ". 2012. "Social networks and spin glasses," *Philosophical Magazine* (92:1-3), pp. 362–377.

Leventhal, T. and Brooks-Gunn, J. 2000. "The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes." *Psychological bulletin* (126:2), p. 309.

Lin, N. 1999. "Social networks and status attainment," *Annual review of sociology*, pp. 467–487.

Ludwig, J. 1999. "Information and inner city educational attainment," *Economics of Education Review* (18:1), pp. 17–30.

Montgomery, J. D. 1991. "Social networks and labor-market outcomes: Toward an economic analysis," *The American economic review*, pp. 1408–1418.

Onnela, J.-P., Arbesman, S., González, M. C., Barabási, A.-L., and Christakis, N. A. 2011. "Geographic constraints on social network groups," *PLoS one* (6:4), e16939.

Phithakkitnukoon, S., Calabrese, F., Smoreda, Z., and Ratti, C. 2011a. "Out of sight out of mind–How our mobile social network changes during migration," in *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on,* IEEE, pp. 515–520.

Phithakkitnukoon, S., Smoreda, Z., and Olivier, P. 2011b. *Socio-geography of human mobility: A study using longitudinal mobile phone data,* Computing Science, Newcastle University.

Rees, A. 1966. "Information networks in labor markets," *The American Economic Review*, pp. 559–566.

Ross, S. L. 1998. "Racial differences in residential and job mobility: Evidence concerning the spatial mismatch hypothesis," *Journal of Urban Economics* (43:1), pp. 112–135.

Sampson, R. J., Morenoff, J. D., and Gannon-Rowley, T. 2002. "Assessing" neighborhood effects": Social processes and new directions in research," *Annual review of sociology*, pp. 443–478.

Song, C., Qu, Z., Blumm, N., and Barabási, A.-L. 2010. "Limits of predictability in human mobility," *Science* (327:5968), pp. 1018–1021.

Soto, V., Frias-Martinez, V., Virseda, J., and Frias-Martinez, E. 2011. "Prediction of socioeconomic levels using cell phone records," in *User Modeling, Adaption and Personalization,* Springer, pp. 377–388.

Stoloff, J. A., Glanville, J. L., and Bienenstock, E. J. 1999. "Women's participation in the labor force: the role of social networks," *Social networks* (21:1), pp. 91–108.

Tizzoni, M., Bajardi, P., Decuyper, A., King, G. K. K., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., and Colizza, V. 2014. "On the use of human mobility proxies for modeling epidemics,"

Toole, J. L., Lin, Y.-R., Muehlegger, E., Shoag, D., González, M. C., and Lazer, D. 2015. "Tracking employment shocks using mobile phone data," *Journal of The Royal Society Interface* (12:107), p. 20150185.

Verkasalo, H. 2009. "Contextual patterns in mobile service usage," *Personal and Ubiquitous Computing* (13:5), pp. 331–342.

Weinberg, B. A., Reagan, P. B., and Yankow, J. J. 2004. "Do neighborhoods affect hours worked? Evidence from longitudinal data," *Journal of Labor Economics* (22:4), pp. 891–924.

Wooldridge, J. 2002. "Econometric analysis of cross section and panel data," *Cambridge, MA: MIT Press. Lower sec. or less* (24), p. 21.

Appendix

	(1) work cell	(2) work cell	(3) work cell
neighbors	0.0013*** (0.0002)	0.0011*** (0.0002)	0.0011*** (0.0002)
friends	0.2998*** (0.0124)	0.2584*** (0.0149)	0.2632*** (0.0155)
both far	-0.0008*** (0.0002)	-0.0008*** (0.0002)	-0.0008*** (0.0002)
friends and both far	0.0747** (0.0284)	0.0675* (0.0282)	0.0417 (0.0370)
neighbors and friends		0.1031*** (0.0229)	0.0912*** (0.0256)
neighbors and friends and both far			0.0563 (0.0569)
intercept	- -	- -	- -
Parish	yes	yes	yes
Observations	1495157	1495157	1495157
R^2	0.027	0.027	0.027

Standard errors in parentheses
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 6. Regression Results - distance between home and work

	friends = I(calls >0)		friends = I(calls reciprocal >0)		friends = text messages >0		friends = I(calls & text messages >0)		friends = I(calls 7pm-7am >0)		friends = I(calls reciprocal >0 & calls 7pm-7am >0)	
	(1) work cell	(2) work cell	(1) work cell	(2) work cell	(1) work cell	(2) work cell	(1) work cell	(2) work cell	(1) work cell	(2) work cell	(1) work cell	(2) work cell
neighbors	0.0013*** (0.0002)	0.0010*** (0.0002)	0.0013*** (0.0002)	0.0010*** (0.0002)	0.0013*** (-0.0002)	0.0011*** (-0.0002)	0.0013*** (0.0002)	0.0011*** (0.0002)	0.0013*** (-0.0002)	0.0011*** (-0.0002)	0.0013*** (0.0002)	0.0011*** (0.0002)
friends	0.2601*** (0.0075)	0.2250*** (0.0087)	0.2884*** (0.0091)	0.2476*** (0.0108)	0.2845*** (-0.0095)	0.2452*** (-0.0115)	0.2899*** (0.0098)	0.2504*** (0.0119)	0.2811*** (-0.0099)	0.2354*** (-0.0119)	0.2936*** (0.0105)	0.2436*** (0.0129)
neighbors and friends		0.1075*** (0.0165)		0.1141*** (0.0193)		0.1056*** (-0.02)		0.1030*** (0.0205)		0.1190*** (-0.0206)		0.1252*** (0.0218)
Parish	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157
R-sq	0.033	0.034	0.030	0.031	0.028	0.028	0.027	0.028	0.026	0.026	0.025	0.026

Standard errors in parentheses
+ p<0.1 * p<0.05 ** p<0.01 *** p<0.001

Table 7. Robustness checks for different definitions of "friendship" with dependent variable "work in the same cell"

	friends = I(calls >0)		friends = I(calls reciprocal >0)		friends = text messages >0		friends = I(calls & text messages >0)		friends = I(calls 7pm-7am >0)		friends = I(calls reciprocal >0 & calls 7pm-7am >0)	
	(1) company	(2) company	(1) company	(2) company	(1) company	(2) company	(1) company	(2) company	(1) company	(2) company	(1) company	(2) company
neighbors	0.0013***	0.0010***	0.0013***	0.0011***	0.0014***	0.0012***	0.0013***	0.0012***	0.0014***	0.0012***	0.0014***	0.0012***
	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
friends	0.6443***	0.6107***	0.7129***	0.6797***	0.7100***	0.6774***	0.7249***	0.6961***	0.7080***	0.6696***	0.7369***	0.7023***
	(0.0081)	(0.0101)	(0.0089)	(0.0115)	(0.0094)	(0.0123)	(0.0095)	(0.0125)	(0.0098)	(0.0130)	(0.0100)	(0.0135)
neighbors and friends		0.1028***		0.0928***		0.0875***		0.0753***		0.0999***		0.0866***
		(0.0167)		(0.018)		(0.0189)		(0.0191)		(0.0195)		(0.0199)
Parish	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157	1495157
R-sq	0.233	0.234	0.209	0.21	0.188	0.189	0.187	0.187	0.174	0.174	0.169	0.169

Standard errors in parentheses
= ” + p<0.1”

Table 8. Robustness checks for different definitions of ”friendship” with dependent variable ”work for same company”