NBER WORKING PAPER SERIES

INFORMATION, MOBILE COMMUNICATION, AND REFERRAL EFFECTS

Panle Jia Barwick
Yanyan Liu
Eleonora Patacchini
Qi Wu

Working Paper 25873
http://www.nber.org/papers/w25873

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2019, Revised October 2021

We thank Susan Athey, Patrick Bayer, Giacomo De Giorgi, Jessie Handbury, Tatsiramos Konstantinos, Mike Lovenheim, Michele Pellizzari, Steve Ross, and various seminar participants for helpful comments. Barwick acknowledges the generous support from the National University of Singapore during her sabbatical visit. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Information, Mobile Communication, and Referral Effects
Panle Jia Barwick, Yanyan Liu, Eleonora Patacchini, and Qi Wu
NBER Working Paper No. 25873
May 2019, Revised October 2021
JEL No. J60,R23

ABSTRACT

We use the universe of de-identified and geocoded cellphone records for 1.8 million individuals from a major Chinese telecommunication provider to examine the role of information exchange in urban labor markets. Information flow, as measured by call volume, correlates strongly with worker flows at different levels of geographic aggregation. Having a referral in a location increases by close to four times the likelihood that a job switcher moves there. Different from the communication pattern between nonreferral pairs, communication between referral pairs exhibits a distinct inverted-U shape that peaks prior to the job switch. Compared with our referral measure, the commonly-used social network proxies deliver a lower bound estimate of the referral effect. We supplement the phone records with administrative data on firm attributes and auxiliary data on job postings and residential housing prices. Referred jobs are associated with higher monetary gains, a higher likelihood to transition from part time to full time, reduced commuting time, and a higher probability of entering desirable jobs. Referral information is more valuable for young workers, people switching jobs from suburbs to the inner city, and those changing their industrial sectors. Firms receiving referred workers are associated with more successful recruits, higher matching rates and faster growth.

Panle Jia Barwick
Department of Economics
Cornell University
462 Uris Hall
Ithaca, NY 14853
and NBER
panle.barwick@cornell.edu

Yanyan Liu
IFPRI
Senior Fellow
Washington, DC 20006
Y.Liu@cgiar.org

Eleonora Patacchini
Department of Economics
Cornell University
430 Uris Hall
Ithaca, NY 14853
and CEPR
ep454@cornell.edu

Qi Wu
Department of Economics
Cornell University
404 Uris Hall
Ithaca, NY 14853
qw98@cornell.edu

1 Introduction

Information affects every aspect of economic decisions, from firm production to household consumption, from government regulation to international treaty negotiations. Classical analysis assumes that agents choose actions to maximize payoff under *perfect* information (Arrow and Debreu, 1954). In reality, information is rarely perfect. Agents' information sets differ substantially, as highlighted by the influential literature on information asymmetry (Akerlof, 1970; Spence, 1973; Rothschild and Stiglitz, 1976). In addition, information exchange and acquisition are costly and crucially depend on social interactions among individuals.

Quantifying the effect of information exchange among social entities and individuals on economic outcomes is challenging because it is difficult to measure the extent of information exchange, and even more so the quality of information that is passed on from one agent to another. The widespread use of location-aware and Global Positioning System (GPS) technologies in mobile phone devices provides a novel avenue that helps researchers to quantify the extent of information flow among individuals, while also tracking their movements in physical space. Datasets derived from these geocoded phone communication records present three unique advantages over traditional ones. First, the frequency and intensity of call records provide a direct measure of information exchange. Second, the panel data nature of these datasets make it feasible to follow individuals over time and space and control for individual unobserved attributes. Third, such data portray a more accurate profile of individuals' social networks than do surveys commonly used in the literature. Existing research has documented that mobile phone usage predicts human mobility (Gonzalez et al., 2008), migration (Blumenstock et al., 2019), poverty and wealth (Blumenstock et al., 2015), credit repayment (Bjorkegren and Grissen, 2018), restaurant choices (Athey et al., 2018), and residential location choices (Buchel et al., 2019).

In this paper, we analyze the impact of information exchange on labor market dynamics. Our analysis is made feasible by a unique dataset that contains the universe of de-identified and geocoded cellphone records from a major Chinese telecommunication service provider over the course of 12 months in a northern Chinese city. These detailed records enable us to construct measures of information flow between geographic areas and among individuals, as well as variables on employment status, history of work locations, home locations, and demographic attributes. We supplement our phone records with administrative data on firm attributes (industry classification and payroll) and auxiliary datasets on residential housing prices and job postings for additional socioeconomic measures.

We first document that information flow as measured by the frequency of phone calls correlates strongly with worker flows. Such a correlation persists at different levels of spatial aggregation. Conditional on the number of phone calls exchanged, the diversity of individuals' social contacts (sources of information) also matters. Within different diversity measures, diversity in socioeconomic status is more valuable than diversity in spatial locations. As far as job mobility is concerned, diversity in the information sources possessed by the working population is far more critical than that by the residential population.

Having illustrated the importance of information flow with respect to worker flows, we examine the role of job-related information shared by social contacts (friends) on job switches.[1] When an individual moves to a preexisting friend's workplace, we define such a friend as "a referral." We first document via event studies that the intensity of information flow between workers and their referrals exhibits an inverted-U shape that peaks at the time of the job switch. In contrast, the information flow between workers and nonreferral friends remains stable throughout the sample period, with no noticeable differences during the months that precede job switches.

We define the referral effect as the effect of having social contacts in a given workplace on individuals' work location choices. We quantify the referral effect using the difference in a job seeker's propensity to switch to a friend's workplace versus a work location in the same neighborhood but without a friend. One might be concerned that our definition of the referral effect suffers from confounding factors (such as firm relocations) or geospatial attributes that are correlated with job flows (commercial centers, industrial clusters, etc.). We tackle these by adding the *interaction* of the origin and destination neighborhood fixed effects. In other words, we compare individuals who share the same origin-destination neighborhood pair but have different social networks and examine their choices of workplace locations with and without friends.

A long-standing challenge in the referral literature that examines observational data is the difficulty in distinguishing a referral effect from homophily and sorting. If individuals share similar skills and preferences with their friends, then an individual might move to a location where a friend works, not because of the referral information but because the vacant position requests certain skill sets. In addition, not all locations have desirable openings. Leveraging the richness and structure of our data, we conduct several falsification tests. First, we limit our analysis to individuals for whom there is at least one additional location within the same neighborhood that has vacancy listings in the same occupation and salary range as the one that the job switcher takes. This mitigates the concern that individuals sort into friends' locations as a result of the availability of suitable job opportunities, rather than useful information provided by referrals. Second, we distinguish between friends currently

[1] We use *social contacts* and *friends* interchangeably in this paper.

working in the location (referrals) and friends who used to work there but moved away prior to the job switch. Given that sorting into friends' locations by unobserved preferences or skills should happen regardless of a friend's *current* location, we would expect to find similar estimates for both types of friends if our estimated referral effects primarily reflect sorting. Third, we compare friends who work in the location (referrals) with friends who live but do not work at the location where a job switcher moves to. Larger estimates for friends working in the location would be consistent with referrals: affiliation with the workplace enables referrals an information advantage of job openings over other types of friends. Lastly, we borrow the "second degree of influence" concept from the network literature (Evtushenko and Kleinberg, 2021) and examine the importance of friends' friends, who are similar to switchers but do not directly communicate with them. Our results from these different tests illustrate that referrals are indeed much more important than friends who moved away prior to the job switch, friends who live but do not work there, or friends' friends, indicating that our estimated referral effects are unlikely to be driven by sorting.

In terms of effect heterogeneity, referrals are particularly important for young workers, people switching jobs from suburbs to the inner city, and those who change sectors. These results are in line with the observation that information asymmetries are more severe in these settings and hence referrals are more valuable. We also provide evidence that strong social ties are associated with a larger referral effect, corroborating results in the literature on weak ties versus strong ties.

Job information passed on via referrals is valuable for workers. Specifically, referral jobs are associated with higher wages and nonwage benefits, shorter commutes, and a greater likelihood to transition from part time to full time and from regular jobs to premium ones. Referral information is more valuable for young workers, people switching jobs from suburbs to the inner city, and those changing their industrial sectors. Information transmitted through the referral networks is also beneficial to firms. Firms receiving referred workers are more likely to have successful recruits, achieve higher matching rates, and experience faster growth. These results provide suggestive evidence that referrals improve labor market efficiency by providing better matches between workers and vacancies and mitigate labor market inequality.

We conducted an extensive set of robustness analyses to examine alternative explanations for the referral effect, including homophily, preference to work with friends, unobserved location attributes, and local labor market demand. We also examined reverse causality, the spatial distribution of friends, as well as alternative friend definitions. Our results survive these additional robustness analyses.

According to our definition, at least one in every four jobs are based on referrals. Having a referral in a location increases by close to four times the likelihood that an individual moves there – a pattern consistent with previous studies carried out in various countries (Ioannides and Loury, 2004). We compare our referral definition with two commonly used measures of social contacts in the literature, namely residential neighbors (Bayer et al. 2008) and individuals belonging to the same ethnic group or immigrant community (Edin et al. 2003), the latter of which is analogous to individuals born in the same county in our setting. We are able to directly validate these measures using call records. As expected, residential neighbors and people born in the same county are more likely to communicate with each other, thus validating these two measures. Nonetheless, while the referral effect from these two alternative referral definitions is positive and significant, it is much lower than our baseline estimate. As a result, the reported estimate in the literature that is derived from these social network proxies is likely to be a lower bound of the true referral effect.

We conclude our analysis with extensions that shed light on the external validity of our findings. Our analysis is based on mobile phone communications because we do not have micro-level data on communication via other channels, such as text messages, mobile apps (such as WeChat), and web-based media (such as emails). Nonetheless, we present evidence that different information channels are complements: people who frequently communicate via phone conversations are also more likely to use other channels, such as text messages and mobile apps, and more likely to browse the Internet. The positive correlation exists both across individuals and within-individuals over time. Furthermore, the referral effect is robust to whether or not individuals' mobile devices are compatible with the 4-G network, a technology that is better for mobile communication apps than earlier generations.

Finally, we replicate our analysis on individuals who experienced unemployment and then successfully found a job during our sample period. This contrasts the bulk of our analysis described above that examines individuals who switch from one job to another with minimal job disruption. Nonetheless, the event studies reveal a remarkably similar pattern in terms of the information flow intensity between unemployed individuals and their referrals. The number of phone calls between these referral pairs also exhibits an inverted-U shape that peaks at the time of reemployment. In addition, the estimated referral effect varies from 0.31 to 0.33, very close to the baseline estimate derived from individuals without an employment gap. While the analysis is constrained to individuals who successfully find a job within a short time window post unemployment, these findings provide suggestive evidence that the communication patterns we document and the referral effect we estimate are potentially applicable to all job seekers, whether employed or unemployed.

Our work contributes to the emerging literature that demonstrates how the widespread use of electronic technologies, and, consequently the wealth of information on individual digital footprints, opens new frontiers for urban economics (Bailey et al., 2018b; Glaeser et al., 2015; Donaldson and Storeygard, 2016). A pioneering study by Henderson et al. (2012) exploits satellite data to conduct an analysis on urban economic activities at a finer level of spatial disaggregation than traditional studies. Using predicted travel time from Google Maps, Akbar et al. (2018) construct city-level vehicular mobility indices for 154 Indian cities and propose new methodologies to improve our understanding of urban development. Other studies examine housing decisions (Bailey et al., 2018a), households' responses to income shocks (Baker, 2018), and entrepreneurship and investment (Jeffers, 2018). Our work contributes to this literature by combining mobile phone records with traditional socioeconomic data to shed light on urban labor market mobility at fine geographical and temporal scales.

Another relevant strand of literature examines the role of social networks in job searches (Topa, 2011; Schmutte, 2016). To identify referred workers, this literature uses surveys or assumes interactions and exchange of job information between social ties, such as former fellow workers (Cingano and Rosolia 2012; Giltz 2017; Saygin et al. 2018), family ties (Kramarz and Skans 2014), individuals who belong to the same immigrant community or ethnic group (Edin et al. 2003; Munshi and Rosenzweig 2013; Beaman 2012; Dustmann et al. 2016; Aslund et al. 2014), residential neighbors (Bayer et al. 2008; Hellerstein et al. 2011; Hellerstein et al. 2014; Schmutte 2015), and Facebook friends (Gee et al. 2017a). The paper closest to ours is Bayer et al. (2008), who also study the importance of referral effects in an urban market. Using Census data on residential and employment locations, they document that individuals who reside in the same city block are more likely to work together than those who live in nearby blocks, and they interpret these findings as evidence of social interactions. We contribute to this literature by providing a more refined measure of social networks and information exchange among individuals, and we introduce complementary data on vacancies and firm attributes to cover a diverse set of economic outcomes.

Our study is also related to the literature on weak vesus strong ties. The seminal study by Granovetter (1973) argues that weak ties could be more important because of their access to a diverse set of information. This study has spurred a large literature on whether weak ties are more effective for information transmission. Aral and Alstyne (2011) show that the importance of weak ties and strong ties could be context dependent. Kramarz and Skans (2014) find that strong social ties are an important determinant of where young workers find their first job. Using Facebook users from 50 countries, Gee et al. (2017b) document that strong ties are more important than weak ties in job finding at the margin, though collectively weak ties are more important because they are numerous. Our results corroborate the findings in the existing literature that the (marginal) referral effect is more pronounced among strong social ties.

Our work is also related to the empirical literature on information economics. Recent studies have shown that increasing information transparency (for example, through better labels and postings) helps consumers' perceptions of product attributes (e.g., Smith and Johnson 1988), improves consumer choices (e.g., Hastings and Weinstein 2008; Barahona et al. 2021), and drives up average product quality (e.g., Jin and Leslie 2003; Bai 2018). Our analysis contributes to this strand of literature by quantifying the importance of information exchange through referrals in facilitating urban labor market mobility. Last but not least, our study is related to the literature on diversity, including Page (2007) and Eagle et al. (2010). We propose novel measures for the diversity of socioeconomic outcomes and illustrate the important role they play in shaping worker flows.

The paper proceeds as follows. Section 2 discusses data, the institutional background, and descriptive evidence. Section 3 presents event studies and baseline regressions on the referral effects. Section 4 analyzes the referral benefits to both workers and firms. Section 5 reports an extensive set of robustness analyses and rules out alternative explanations. Section 6 examines our finding's external validity and compares our referral measure with proxies for social interactions that are commonly used in the literature. Section 7 concludes.

2 Data and Descriptive Evidence

2.1 Data

We compiled a large number of datasets for our analysis. Besides data on geocoded phone records, we assembled administrative data on firm attributes and auxiliary data on neighborhood attributes, residential housing prices, and vacancies (job postings).

Geographical Units At the highest level, the city we study is divided into 23 administrative districts and counties.[2] These districts and counties are further broken into 1,406 neighborhoods that are delineated by major roads. A neighborhood is similar to but smaller in size than a census block in the United States. There are 917 neighborhoods in the urban center of the city and 489 neighborhoods in surrounding suburbs.[3] The lowest level of a geographical unit is a *location*, a geographic position returned by a cellular tower station, which

[2]The city consists of an urban core divided into 8 districts, and 15 surrounding suburban and rural counties. These 8 districts and 15 counties are all equal parts of the city proper and under its administrative authority.

[3]These neighborhoods are constructed by our data provider for billing purposes. The average size of an administrative district/county, a neighborhood in the urban core, and a neighborhood in the suburb is 712.00 km^2, 0.45 km^2, and 25.03 km^2, respectively. See Online Appendix Figure S1 for a section of the city map.

represents a building complex or an establishment within a neighborhood. The median and average number of distinct locations in a neighborhood is 7 and 13, respectively. In total there are close to 18,000 locations.

Spatial attributes come from two GIS shape files (maps). The first shape file delineates administrative divisions, roads, highways, railways, and parks, as well as points of interests, such as hospitals, schools, shopping malls, parking lots, and restaurants. The second shape file depicts neighborhood boundaries. We overlay these shape files to obtain the spatial attributes for each location and neighborhood.

Call Data China's cellphone penetration rate is very high. According to the China Family Panel Studies (CFPS), a nationally representative longitudinal survey of individuals' social and economic status since 2010, 85 percent of correspondents sixteen years and older report possessing a cellphone.

Our anonymized and geocoded call data contain the universe of phone records for all mobile phone subscribers (excluding commercial entities) of a major Chinese telecommunications company in a city, covering the period of November 2016 to October 2017. The data provider (hereafter Company A) serves between 30–65 percent of all mobile phone users in the city we study.[4]

Cellphone usage records are automatically collected when individuals send a text message, make a call, use apps, or browse the Internet. These records include individual identifiers (IDs), location at the time of usage, and the time and duration of usage. The data we have access to are aggregated to the weekly level and contain encrypted IDs of the calling party and the receiving party, call frequency and duration in seconds, whether or not a user is Company A's subscriber, and demographic information about subscribers, such as age, gender, and place of birth. The birth county enables us to distinguish migrants from local residents. The existing literature has shown that migrants are much more likely to refer and work with other migrants from their birth city and province (Dai et al., 2018).

An important advantage of our data is that the geocoded location is recorded whenever the mobile device is used and every 15 minutes when the device is turned on. The serving cellular tower station records a geographic position in longitude and latitude that is accurate up to a 100–200 meter radius, or roughly the size of a large building complex. For each individual and week, we observe the location that has the most frequent phone usage (calls,

[4]China has three major telecommunications companies. We report a range for the market share to keep Company A anonymous. For individuals with multiple phones, we observe usage on the most commonly used phone. If they subscribe to services from multiple carriers (which is uncommon), we only observe activities within Company A. China adopted the "real-name system" in 2011; since January 2017, mobile phones that do not pass the real-name authentication cannot operate in China. This allows us to identify individuals based on their anonymized IDs.

texts, apps, Internet browsing, etc.) between 9 am and 6 pm during the weekdays (which we call a "work location") as well as the location that has the most frequent usage between 10 pm and 7 am for the same week (which we call a "residential location").[5] In contrast to traditional datasets in social science studies that lack fine-grained geographical information about human interactions, these geocoded locations trace out individuals' spatial trajectories over time and allow us to construct a diverse set of social ties (including friends, neighbors, past and present coworkers, friends' coworkers, etc.).

Constructing individuals' workplace history using recorded geocodes is the most crucial step of our analysis. Since we do not directly observe the employment status or place of work, we take a very conservative approach to mitigate measurement errors in work-related variables. The raw data contain 1.8 million individuals. We focus on those with valid work locations for at least 45 weeks – a period long enough to precisely identify workplaces. Locations that are visited during the working hours on a daily basis for weeks in a row are likely to be a workplace rather than shopping centers or recreational facilities. This gives us 560,000 individuals.[6] After further restricting to individuals who have at most two working locations throughout the sample period (which excludes sales persons and individuals with out-of-town business travels and family visits) and for whom we have the complete demographic information, our final sample reduces to 456,000 users. We carry out the core empirical analysis using this sample and conduct robustness checks in Section 5 using less stringent sample selection criteria.

Job Switchers We identify individual i as a *job switcher* if the following criteria are satisfied.[7] First, a job switcher is someone who worked in two work locations, is observed at least four weeks in either location, and switches locations only once. Second, the distance between these two locations must be at least 1 kilometer (km). We choose the cut-off of 1 km to avoid erroneously identifying someone as a switcher, because individuals' work locations are geocoded up to a radius of 100-200 meters (the average distance between neighborhood centroids is 1.4 km). Among the 456,000 users in our final sample, 8 percent (38,102) are identified as job switchers. Though constructed using different data sources, this on-the-job switching rate is similar to that reported in the literature for China's labor market, which is around 7 percent (Nie and Sousa-Poza, 2017). China's job-to-job mobility is lower than that

[5] Phone usage during 7 am–9 am and 6 pm–10 pm is excluded because people are likely on the move during these time intervals.

[6] Several factors contribute to the sample attrition. China's cellphone market is dynamic, with a high fraction of subscribers switching carriers during each month, especially among people on prepaid plans. In addition, the work location information is missing for weeks when individuals travel out of town or experience frequent location changes (common for unemployed or part-time workers, salesmen, etc.).

[7] Online Appendix Figure S2 depicts the timeline of job switches.

in Western countries (for example, 15-18 percent in the European Union, as documented in Recchi 2009), partly because of the hukou system, which imposes significant restrictions on individuals' migration across provinces or from rural to urban areas (Ngai et al., 2017; Whalley and Zhang, 2007). Our switchers found jobs in a total of 5,800 unique work locations that are spread across in 1,100 neighborhoods. Two-thirds of these locations are in the urban core; the remainder are in surrounding counties.

Vacancy Data To gauge the dynamics of local labor market conditions, we collect listings from the two largest online job posting websites, Zhilian (zhaopin.com) and 58.com, from August 2016 to February 2018.[8] These websites hold on average 10,000 job postings per month. We obtained a total of 121,055 postings and merged them with our call data based on locations.

Each posting reports the posting date, job title and description, full time or part time, qualifications (minimum education level and years of experience), monthly salary (in a range), firm address, firm size (number of total employees), and firm industry. On the basis of the job title and description, we group these postings into eight occupations using the 2010 U.S. occupation code (see Online Appendix S1 for more details). Popular occupations include Professionals (26.70 percent), Service (26.61 percent), Sales and Office Administration (19.24 percent), and Management (17.47 percent), followed by Education, Legal, Arts and Media (11.53 percent), Farming, Fishing, and Construction (6.44 percent), Production and Transportation (2.29 percent), and Health Related (1.45 percent).

The vacancy postings report a salary range (for example, an annual salary of RMB 25,000–40,000) instead of the exact job compensation. In practice, once the job is taken, a sizable fraction of the worker compensation consists of nonwage benefits, including bonuses and commissions, paid vacations, and health and unemployment insurance, etc.(Cai et al., 2011). For these reasons, we rely on the payroll information from the firm administrative data (see below) to measure job compensation.

Administrative Firm-Level Records We use two firm-level administrative datasets to obtain wages and benefits, local industry composition, and firm attributes. The first is the annual National Enterprise Income Tax Records from 2010 to 2015, which is collected by the State Administration of Taxation and contains firm ID, industry, ownership, balance sheet information (revenue, payroll, employee size, etc.), and tax payments. This database

[8]Zhilian.com reported a 27.5 percent market share in the fourth quarter of 2017 and became the largest online posting platform in the second quarter of 2018 (https://www.analysys.cn/article/detail/20018775). The website 58.com is a close second, accounting for 26.5 percent of the market in the fourth quarter of 2017 and serving more than 4 million firms (http://www.ebrun.com/20161230/208984.shtml).

oversamples large companies (major tax payers) and small-sized firms and undersamples medium-sized firms, covering about 85-90 percent of the city's GDP. Location information is obtained by merging these tax records with the Business Registration Database that is maintained by China's State Administration for Industry and Commerce. Our final dataset contains firm location, industry, ownership type (whether or not state owned), employee size, revenue, wage payroll, and capital, for a total of between 5,000 to 10,000 firms.[9]

In our sample, most firms are private (85.6 percent), followed by state-owned (7.0 percent), foreign (0.7 percent), and other ownership types (6.6 percent). Over 60 percent of firms belong to the manufacturing sector, which is higher than the national average of 25.4 percent (National Bureau of Statistics of China, 2014) and reflects the industrial focus of the city. Using the average payroll as a measure of job compensation, jobs in nonmanufacturing firms are paid significantly higher than those in manufacturing firms, demanding nearly a 50-percent premium (an average annual wage of RMB 32,005 vesus RMB 20,609).

We assign each firm in the tax data to the nearest location reported in the geocoded call records and cap the distance at 500 meters. Firms that are farther away are dropped. For 79 percent of job switchers, job compensation is obtained from the payroll of a firm within 300 meters. For locations with multiple firms, we use the employment-weighted payroll to more accurately reflect an average worker's compensation in a location.

Housing Price Our main data source does not contain individuals' socioeconomic measures such as wealth or income. To overcome this data limitation, we scrape housing data from Anjuke.com, a major online real estate brokerage intermediary and rental service provider in China that collects housing information for both residential and commercial properties. For each residential complex, Anjuke.com reports its name, property type and attributes, the monthly average housing price per square meter, year built, total number of units, average size, and street address. We successfully merged 64 percent of the neighborhoods in the urban core (city center) and 20 percent of neighborhoods in surrounding counties with residential neighborhoods in Anjuke.com.

These data sources provide information on a large number of attributes for each location and neighborhood, including the most common occupations among job postings, industry composition, number of employees and vacancies, average wage, and housing price. For each individual in our final sample, we observe his/her work and residential location, friends, neighbors, as well as the workplaces and home locations for both friends and neighbors.

[9] The exact number of firms is omitted to keep the city anonymous.

2.2 China's Labor Market and Referrals

China's labor market has several noticeable features. Relative to other developing countries, China has a high female labor participation rate. In response to the employment pressure generated by its large population, China has instituted a mandated early retirement age of 55 for female workers and 60 for male workers.

Established in the 1950s, China's hukou system categorizes individuals as agricultural or nonagricultural on the basis of their birth place, partly to anchor peasants to the countryside. According to Zhang and Wu (2018), China's labor market has a two-tier system: urban cities and rural areas. The large divide that separates these two tiers in terms of job opportunities, social benefits, and amenities (education, health care, etc.) has created a large number of migrant workers in urban cities who take jobs with low wages and long working hours, and who are often denied social benefits.

Similar to the United States and European countries, referrals are common among Chinese workers. Figure 1 compares the popularity of different job search methods among Chinese and U.S. workers using data from the 2014 CFPS and the 2014 U.S. Current Population Survey, respectively. Workers in China (red dotted bars) are more likely to rely on informal search methods (38 percent of workers in China find jobs through friends, compared to 30 percent in the United States), while formal search methods, such as ads, job agencies, or contacting employers directly, are more prevalent in the United States (blue solid bars). In addition, referral is more important for young workers in China, with a higher fraction of young correspondents citing referrals as their main channel of finding a job.

The top panel of Table 1 presents descriptive statistics of all individuals in our sample. About one-third (36 percent) of users are women and 89 percent of users are younger than 60, reflecting the higher mobile-phone penetration among men and the younger population. Three-quarters of our sample users were born in the local province; the rest migrated from other provinces. Thirty-nine percent of users were born in the city we study. The last column presents the national average of the 2014 CFPS survey among individuals who use a cellphone.[10] Our sample exhibits similar demographics as the national average, except it contains a smaller fraction of those under age 25 and fewer women, partly because we focus on individuals with stable jobs.

Social Ties and Referrals The bulk of our analysis focuses on job switchers and their social ties. As the bottom panel of Table 1 illustrates, job switchers bear similar demographics as non-switchers, except for age. Job switchers are more likely to be in their thirties and on

[10]The CFPS sample is restricted to adults with phone-related expenses that exceed RMB 30 per month to ensure proper phone usage.

average two years younger than nonswitchers. They are more likely to be migrants and have a higher fraction of friends who use Company A's mobile service, although the magnitude of these differences is modest.

Switcher i's social contacts include everyone who makes a phone call to or receives a phone call from individual i at least once during our sample period.[11] The call data provide rich information on users' social network, but only contain information on work locations for Company A's subscribers. On average, 50 percent of a user's friends are Company A's subscribers. One might be concerned about the potential sample selection bias if Company A's subscriber network overrepresents certain demographic groups. This is unlikely to be a major concern. First, Company A's network of users is geographically spread out and covers all street-blocks of the city. Second, pricing and plan offerings are similar across mobile service providers. Nonetheless, to examine the robustness of our results, Section 5 separates individuals whose friend coverage is above the median from the rest and documents similar findings.

Figure 2 reports the average weekly number of social contacts that switchers communicated with over our sample period. It varies between 23 and 25 for most weeks. Importantly, no spikes arise in the number of social contacts during the weeks leading to the job switch. Instead, the number of friends decreases modestly prior to the job switch and becomes slightly higher after the switch.[12] These patterns suggest that social links that are established prior to the job change are likely exogenous; otherwise we should expect a spike in the number of contacts during weeks approaching the job change. Nonetheless, to mitigate concerns of endogenous links that are formed surrounding the job change, we use social contacts established *three months* prior to the job switch throughout the empirical analysis (except when noted otherwise).[13] Section 5 documents robustness to alternative cutoffs (excluding social contacts formed within one month, two months, ..., to five months of the job switch.)

When a switcher moves to a preexisting social contact's workplace, we define such a friend as a "referral". Among the 38,102 job switchers observed in our sample, friend location information is missing for 4,703 (12 percent) of them (Panel A of Table S1). Among the switchers with nonmissing locations for at least one friend, 25 percent find a job through a referral. Note that this should be interpreted as a lower bound as we limit to friends with 45 weeks of nonmissing work locations. As discussed in Section 2.1, 45 weeks ensures the

[11] These are called "one-way" contacts. An alternative definition requires a contact to both make a phone call to *and* receive a phone call from individual i at least once during the sample period. Robustness analyses in Section 5 illustrate that these two definitions lead to very similar results.

[12] Observe that commercial entities are excluded from our data source; see Section 2.1.

[13] This is also consistent with the observation that Chinese employers' recruitment decisions are often made one to two months before the expected start date, due to China's Labor Contract Law, which requires a 30-day notice before the termination of an employment contract.

accuracy of identified job switches, but it may underreport the fraction of referred job moves. In Panel B of Table S1, we relax the friend sample to include all social contacts with at least four weeks of nonmissing work locations. Among switchers with friend location information, 43 percent move to a referral friend. In light of this difference, Section 3 presents estimates with our preferred friend definition, while Section 5 repeats these analyses using friends with at least four weeks of work information.

2.3 Motivating Evidence

Before we delve into a formal regression analysis, we present motivating evidence that a strong correlation exists between information flow (proxied by the intensity of phone communications) and worker flows. In addition, information diversity also matters and the information diversity of the working population has a much stronger correlation with worker inflows than that of residents in the same location. To the best of our knowledge, this is the first analysis that examines the empirical relationship between information flow and worker flows.

Information Flow and Worker Flows We use the number of phone calls between two geographic areas to measure the information flow.[14] Worker flow is defined as the total number of job switches between two geographic areas.

There is a strong correlation between the information exchange (phone call volume) and worker flows. For example, the raw correlation between these two series at the city district level exceeds 0.94. Some correlation arises naturally from heterogeneous spatial and economic attributes, such as economic centers and urban cores. To illustrate the patterns of worker flows and mobile communication, we run a regression analysis and control for origin and destination fixed effects (top panel of Table 2). We use three different levels of spatial aggregation – district pairs (Columns 1 and 2), neighborhood pairs (Columns 3 and 4), and location pairs (Columns 5 and 6). All regressions include two-way area fixed effects to control for differences in economic activities and spatial attributes.

When the two-way district fixed effects are the only regressors (Column 1), the R-squared is 0.745, suggesting that district fixed effects alone can explain three-quarters of variation in worker flows across district pairs. The R-squared jumps to 0.971 when the total number of phone calls between a district pair is included as an additional regressor (Column 2), suggesting that information flow has a strong correlation with worker flows. The effect size is both statistically and economically significant: 300 more calls are associated with one more

[14] An alternative measure of information flow, the total call volume in minutes, delivers similar results.

job switch. A log-log specification suggests that doubling the number of calls is associated with a 35 percent increase in worker flows.

The strong correlation persists at finer geographical areas. There are a total of 987,713 neighborhood pairs and 159,856,138 location pairs. Explaining worker flows at these disaggregated geographical levels is a demanding exercise. Reassuringly, the positive correlation exists even at these fine scales, with 5,000 calls associated with one additional worker flow at the neighborhood-pair level (Column 4). To better measure job-related information that facilitates worker flows, we separate calls by nonswitchers from those by switchers, and then limit to calls received or made by job switchers prior to their job change, one month prior to job change, and three months prior to job change (Online Appendix Table S2). The patterns are qualitatively similar to those in Table 2, except that the coefficient increases as we limit to calls by switchers, as expected.

Existing studies have shown that mobile phone usage can predict economic activities (Kreindler and Miyauchi, 2019). Panel (b) of Table 2 illustrates this point. We regress worker flows on two-way area fixed effects and call volumes using the first six months of the sample, predict worker flows for the second six months, and report the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) for the prediction exercise. Including call volume as a regressor significantly improves prediction accuracy in all cases. For example, RMSE reduces from 66.6 calls to 26.3 calls and MAPE shrinks from 596.7 percent to 115.3 percent when call volume is included for the district pair analysis.[15]

In summary, call volume, which proxies for job-related information flow, has a high correlation with worker flows and can greatly improve prediction accuracy at all levels of spatial aggregation.

Social Tie Diversity and Economic Outcome Both the sociology and economics literature have long emphasized the importance of diversity (Ottaviano and Peri, 2006; Ashraf and Galor, 2011; Alesina et al., 2016). In our setting, the content and value of information might vary over time and across individuals. A high volume of information exchange that is limited to the same area or social group might not be as beneficial as information from a more diverse setting that taps into different social entities.

Following Eagle et al. (2010), we define three diversity measures using the normalized Shannon entropy: social entropy, spatial entropy, and income entropy.[16] These entropy measures reflect the complexity of an individual's network in terms of socioeconomic status and spatial coverage. We average the diversity measures over all individuals who reside or

[15] The large MAPE is partially driven by significant differences across district pairs.

[16] The Online Appendix S2 presents precise definitions of these measures. Further details on information theory and entropy measures can be found in Cover and Thomas (2006).

work in each location. A high value indicates that the working or residential population at a particular location communicates with a diverse source of information. To examine the importance of diversity, we regress the log of worker flows into a given location on the average entropy measures at the location level. Our controls include the total call volumes, which, as shown above, are an important predictor of worker flows; the number of individuals (that is, subscribers of Company A) observed in a location, which captures the scale effect (more populated areas naturally have a higher job inflow); and neighborhood fixed effects. The standard errors are clustered by neighborhood.

Table 3 reports the results. Columns 1 to 3 include one entropy measure at a time, while Column 4 stacks all three measures together.[17] Social entropy, and especially income entropy, which reflects the socioeconomic diversity of individuals' information sources, has a sizable and significant impact on job inflow conditional on the total number of calls. A one standard deviation increase in social and income entropy is associated with a 3 percent and 10 percent increase in worker inflow, respectively. In contrast, the coefficient of spatial entropy is insignificant, partly because our sample consists of individuals from the same city with limited spatial diversity. Online Appendix Table S3 examines the relative importance of information possessed by the working versus residential population and illustrates that the information diversity of the working population has a much stronger correlation with worker inflows than that of residents in the same location. While our analysis is descriptive, these results highlight the heterogeneous values of information possessed by different social groups and reflect the fact that information about jobs exists predominantly in the domain of the working population.[18]

Having illustrated the high correlation between information exchange and job flows, we now turn to the bulk of our empirical analysis, which focuses on a specific channel of information at work: information on job openings shared among social contacts. The existing literature has documented that 30 to 60 percent of all jobs are typically found through informal contacts rather than formal search methods, a pattern that holds across countries and over time, regardless of occupation or industry (Topa, 2001; Burks et al., 2015). We next use our call data to depict individuals' social network and quantify the magnitude of referral

[17] Here we limit to locations that have at least 5 workers and 5 residents. Results are similar if we use all locations or limit to those with at least 10 observed workers and 10 residents.

[18] It is worth noting that our results are remarkably similar to the findings in Eagle et al. (2010), who examine U.K. phone calls in 2005 and relate communication flows to communities' socioeconomic well-being. While the average number of monthly contacts is higher in our context (24 versus 10.1 in the United Kingdom, which reflects a denser social network in China), the average minimum number of direct or indirect edges that connect two individuals is very similar (10.4 in our context versus 9.4 in the United Kingdom). As in our setting, a strong correlation (varying from 0.58 to 0.73) exists between information diversity and the socioeconomic development of communities in the United Kingdom. These results indicate that information diversity is at play across different socioeconomic contexts and not limited to specific regions or time periods.

effects as well as the referral benefits to workers and firms.

3 Communication and Job Changes

In this section, we first conduct an event study to examine the time series variation of the information flow between job seekers and their referrals versus the information flow between job seekers and nonreferral friends. We then perform regression analyses to quantify the referral effect and explore effect heterogeneity.

3.1 Event Study

The detailed call records allow us to examine communication patterns between a job switcher i and his referral versus nonreferral friends over time. In contrast to the stable number of social contacts as shown in Figure 2, the call frequency between switchers and their contacts exhibits interesting dynamics. Using an event study, we examine the call frequency dynamics during the event window of 11 months before and 9 months after the job switch with a rich set of fixed effects:[19]

$$\text{Freq}_{ijt} = \sum_{s=-11}^{9} \gamma_s \text{Referral}_{ij} \cdot 1[t = s] + \sum_{s=-11, s\neq -1}^{9} b_s \text{nonreferral}_{ij} \cdot 1[t = s] + \lambda_i + \tau_t + \epsilon_{ijt}, \quad (1)$$

where Freq_{ijt} is the number of calls between switcher i and his friend j in month t. Referral_{ij} takes value one if switcher i moves to friend j's workplace during the sample period and zero otherwise. nonreferral_{ij} takes value one for all other friends. Note that friend types do not vary over time by construction. The key coefficients $\{\gamma_s, b_s\}$ vary by event month s ($s = 0$ for the month of job switch). The call frequency between a switcher and nonreferral friends before the job switch at time $s = -1$ is the reference category. All regressions include individual fixed effects λ_i, which control for personality such as out-going or introverted personal traits, and month fixed effects τ_t, which control for holidays and seasonality. With individual fixed effects, the event study coefficients γ_s and b_s do not reflect the level of call frequency between friend pairs. Instead, they capture dynamic changes in call frequency *relative to* an individual's baseline talking frequency to nonreferral friends prior to the job switch.

Figure 3 presents the results. To facilitate identification of the event study coefficients and

[19]Since we need a three-month window prior to the job change to define switchers' social network, the event window after the job change has a maximum of nine months. We use a monthly event window instead of a weekly window to average out noises in time trends.

to increase the underlying sample size for the tail months (Schmidheiny and Siegloch, 2020), the period prior to the ninth month before the job switch is binned with the ninth month, and the ninth month post the job switch is binned with the eighth month. The standard errors are clustered at the individual level, though results are robust if we cluster by the work neighborhood instead. The confidence intervals for nonreferral friends are much tighter than those for referral friends. This is because switcher-nonreferral pairs are more common: there are 4.76 million switcher-nonreferral-month observations relative to 238 thousand switcher-referral-month observations.

The communication patterns between referral pairs and nonreferral pairs are distinct, even after controlling for a rich set of fixed effects. First, switchers have more frequent calls with referral friends than with other friends. This pattern corroborates findings from the literature (Gee et al., 2017b) that referral friends are closer social contacts. Second, the intensity of information flow between referral pairs exhibits an inverted-U shape that peaks just before the job change. In contrast, the information flow between switchers and other friends remains stable throughout the sample period, with no noticeable change in the months prior to the job switch. Lastly, the communication intensity between referrals and referees remains elevated post the job switch. Information flow appears to increase with the dimensions of social interaction, as referrals and referees are friends before the job switch and become friends and colleagues afterward.

One might be concerned that individuals sometimes share news about their job offer with friends, which would also lead to intensified communication before they move to the new workplace. However, if this were true, we should expect to observe a spike in the communication volume with *both* referral *and* nonreferral friends. The fact that we do not see such an increase with nonreferral friends indicates that the communication between workers and referrals is unlikely to be driven by workers informing friends of their job change. Finally, some phone calls between the referral pairs could be inquiries about workplace amenities (instead of job openings per se). We regard all such calls as communication with referrals that facilitates a job switch.

3.2 Regression Analysis

We now turn to a regression framework to quantify the referral effect that shapes job seekers' location choices. We limit our analysis to job switchers.[20] Specifically, we compare the

[20] Section 6 reports results for individuals who experienced unemployment gaps and then subsequently found a new job during our sample period.

propensity for switcher i to find a job at a friend's workplace versus a nearby location:

$$M_{il} = \beta \text{Friend}_{il} + \boldsymbol{X_i Z_l \gamma} + \varepsilon_{il}$$

where M_{il} is one if i moves to location l. We restrict individual i's choices to locations *within* the neighborhood that contains his new workplace. This is done purposefully. Job location choices are influenced by many factors, including industry composition and labor demand, commuting distance and local amenities, and intra-household bargaining for married couples, many of which are unobserved in our setting. Limiting individuals' choices to locations within the neighborhood of their new workplace greatly reduces the extent of heterogeneity across choices and allows us to better isolate the effect of referrals from competing explanations of job changes. One implication of this regression design is that demographic variables (or individual fixed effects) do not help explain location choices as they are invariant across locations.

The key regressor is $Friend_{il}$, a dummy variable for having at least one friend working in location l. We include a rich set of interactions between demographic attributes and location amenities. The demographic controls ($\boldsymbol{X_i}$) consist of a constant, gender, migrant status, and age group categories (age 25-34, age 35-44, age 45-59, 60 and above). We also include i's total number of social contacts (irrespective of carriers) to capture differences in personality and social outreach. Amenities at each work location ($\boldsymbol{Z_l}$) are captured by the number of restaurants, the number of roads and parking lots (which measures a location's accessibility), as well as the number of schools within a 500-meter radius. To allow for differential preferences toward local amenities, we interact gender with schools and parking lots, age group dummies with restaurants, the migrant dummy with the number of roads, and the number of i's social contacts with all location attributes. Results are similar if we control for the full interactions between all demographic characteristics and location attributes. To help interpret the magnitude of the coefficient estimates, Table S4 tabulates summary statistics for key variables referenced in various regression samples.

The coefficient of interest is β, which captures the referral effect. There are two main threats to a causal interpretation. The first threat arises from unobserved spatial confounders where a positive correlation can arise with exogenous worker flows from one area to another (for example, firm relocation). In addition, there are unobserved location attributes that affect worker flows across areas. We tackle this problem by adding the *interaction* of the origin and destination neighborhood fixed effects:

$$M_{il} = \beta \text{Friend}_{il} + \boldsymbol{X_i Z_l \gamma} + \lambda_{\tilde{c},c} + \varepsilon_{il} \tag{2}$$

where $\lambda_{\tilde{c},c}$ is a dummy for the pair of neighborhoods ($\tilde{c}$,c) that contains individual i's previous and current workplace. There are a total of 21,000 neighborhood-pair fixed effects in the full sample. This is a demanding specification wherein the key coefficient β is estimated from the within origin-destination variation. We essentially compare individuals who move between the same origin-destination neighborhood pair but have different friend networks and examine their location choices.

The second long-standing challenge in the literature using observational data is the difficulty in distinguishing a referral effect from homophily and sorting. If individuals share similar preferences and skills with their friends, then a positive β could be driven by sorting rather than referrals. In addition, not all locations have desirable openings. An individual might move to location l not because of referrals but because other locations lack appropriate job opportunities. In other words, the friend dummy might simply proxy for locations specializing in jobs that require similar skills shared by individuals and their friends.

Leveraging the richness and structure of our data, we propose the following battery of tests. First, we limit our analysis to workers for whom there is at least one other location within the same neighborhood that has vacancy listings in the same occupation and offering the same salary range as the job taken.[21] This mitigates the concern that individuals sort into friends' locations that provide the only employment opportunity in the area.

Second, we distinguish between friends who are currently working in location l and friends who used to work there but moved away prior to the job switch. Given that sorting by unobserved preferences or skills should happen regardless of a friend's *current* location, we would expect to find similar β estimates for both types of friends if our finding is driven by sorting.[22] Third, we distinguish between friends who work versus friends who live at location l. Larger estimates for friends who work in location l would be consistent with the referral effect: affiliation with the workplace enables friends who work there to have an information advantage of jobs openings. Lastly, we borrow the "second degree of influence" concept from the network literature (Evtushenko and Kleinberg, 2021) and examine the importance of friends' friends. These second-degree links are similar to switchers but do not directly communicate with switchers by construction.

Results Table 4 reports the coefficient estimates for model (2). All columns include the old-by-new neighborhood-pair fixed effects and interactions of demographic and location

[21] The occupation of location l is the most common occupation among all postings. It is coded as missing if the most common occupation accounts for less than 33 percent of all postings at the same location. Results are robust if we use the expected job compensation as measured by the average payroll (see Section 4.1) instead of salary range.

[22] This exercise assumes that homophily is time-invariant, which is likely to hold during our short sample period.

attributes.[23] The standard errors are clustered at the neighborhood-pair level, though the statistical significance of key parameter estimates is robust to the choice of clusters. The mean propensity to choose a location within a neighborhood is 0.09. The referral coefficient in Column 1 is 0.34. It is economically large and precisely estimated. The probability of moving to location l increases by close to four times with a friend working there.

Column 2 limits to the subset of switchers who have at least one alternative work location within the same neighborhood that has openings in the same occupation and offers the same salary range as the one they take. This exercise speaks to the concern that having a friend in a workplace simply proxies for "jobs that require similar skills shared by individual i and his friends." If this were true, whether or not there is a friend in a workplace should not matter as much for individuals facing multiple job opportunities that are similar. Instead,in Column 2 we continue to see a positive and large effect of having a friend. Note that this robustness analysis is made possible by the rich information on vacancies we collected. It mitigates the concern that individuals sort into friends' locations because these are the only suitable employment opportunities in the area.

As it is important to control for the availability of job openings, all regressions hereafter use this subset of switchers who face multiple similar job opportunities. This only moderately reduces the number of observations from 1,120,797 to 915,251. In subsequent discussions, we refer to Column 2's estimate of the referral effect (0.35) as the baseline estimate.

Column 3 contrasts friends currently in the new work location with friends who recently moved away, while Column 4 compares friends working versus friends living in the new workplace. In both cases, friends currently working in the new location have a much larger impact on the choice probability: they are five times as influential as friends who recently moved away and twice as effective as friends who live but do not work in the same location. The differences in parameter estimates are statistically significant at the 1 percent level.

Column 5 compares referral friends with second-degree links – friends' friends – who also work in location l. These individuals reflect homophily because as "friends' friends," they are similar to individual i. On the other hand, they are not friends with switcher i and hence are unlikely to communicate job information with switcher i. The coefficient of friends' friends is only 0.14, much smaller than the coefficient of referral friends (0.34), which remains stable across columns.

To provide more direct evidence for these falsification tests, we repeat the event studies and examine the communication patterns with different types of friends. We present two event study figures as the graph with four types of friends is hard to read, though results

[23] Table S5 presents supplementary results for Column 1 that begin with no controls and increasingly add more regressors and fixed effects.

are very similar if we pool all friend types in one regression:

$$\text{Freq}_{ijt} = \sum_{s=-11}^{9} \gamma_s \text{Referral}_{ij} \cdot 1[t=s] + \sum_{s=-11}^{9} \alpha_s \text{MovedAway}_{ij} \cdot 1[t=s] + \sum_{s=-11, s\neq -1}^{9} b_s \text{OtherFriends}_{ij} \cdot 1[t=s] + \lambda_i + \tau_t + \epsilon_{ijt},$$

and

$$\text{Freq}_{ijt} = \sum_{s=-11}^{9} \gamma_s \text{Referral}_{ij} \cdot 1[t=s] + \sum_{s=-11}^{9} \alpha_s \text{LiveAtNewPlace}_{ij} \cdot 1[t=s] + \sum_{s=-11, s\neq -1}^{9} b_s \text{OtherFriends}_{ij} \cdot 1[t=s] + \lambda_i + \tau_t + \epsilon_{ijt},$$

where MovedAway_{ij} flags friends who moved away before switcher i joined the new work location, and $\text{LiveAtNewPlace}_{ij}$ takes value one if friend j lives in the neighborhood that contains switcher i's new work location but does not work at the new job location. These two event studies are shown in Figure S3 and Figure S4. Echoing findings in Figure 3, the communication patterns between the referral pairs that peak immediately before the job switch exhibit a sharp contrast to those between other types of friends. Specifically, the call frequency between other types of friends is flat during the months surrounding the job switch, indicating that information exchanged between these friend pairs is unlikely to be specific to the job change. The communication pattern with friends who moved away is somewhat noisy due to the limited number of observations, though overall there is no evidence that it peaks close to the job switch. These pictures provide direct evidence that the increased communication intensity with switchers is specific to friends who work in the new workplace. While friends living in the new workplace's neighborhood might have relevant information on local amenities and friends who used to work in the new workplace might share similar skills and preferences, there is no systematic evidence that they provide job-specific information, in contrast to referral friends.

The referral effect is economically large, precisely estimated, and stable across all columns in Table 4. In addition, the communications between referral friends exhibit remarkably different dynamic patterns from those between other types of friends. These results cannot be reconciled with sorting and indicate that referrals at work carry useful information that facilitates the matching between workers and job openings.

Effect Heterogeneity Referrals could facilitate the match between job seekers and vacancies in different ways. For example, current employees can share job opportunities with their social contacts (information to workers). Alternatively, employees can inform their employer of their friends' work attitude and labor market prospects (information to firms). Although we cannot disentangle these different mechanisms, we test their common implication that referrals mitigate information frictions in the hiring process. We thus examine whether referrals are more important when information asymmetry is more severe.

Individuals who live far away from the new work location, have limited work experience, or change industrial sectors are likely to be disadvantaged when it comes to obtaining information about new job openings. Similarly, employers are less likely to be knowledgeable about these workers. In Table 5, we interact Friend_{il} with the distance between the old and new workplace, the distance between home and the new work location, a dummy for young workers (between 25 and 34), a dummy for moving from rural to urban locations, and a dummy for changing sectors. Referrals facilitate job transitions in *all* these situations, especially for rural workers migrating to urban areas and for people changing industrial sectors. For these two groups of individuals, the point estimate of the referral effect is 0.66 and 0.53, respectively, a significant boost above the base estimate of 0.35.

The evidence in Table 5 also helps rule out a couple of alternative explanations. One is that our results are simply driven by preferences: individuals enjoy the company of friends and hence prefer to work at their place. However, this theory cannot explain the stronger referral effect when information asymmetry is more severe (or the communication patterns documented in Section 3.1).[24] Similarly, could our estimates be driven by nepotism, that is, friends and family are hired instead of the best available candidates (Hoffman, 2017)? This is probably not first-order since this theory cannot predict that the effect is stronger when information asymmetry is more severe. Moreover, as shown in Online Appendix Section S3 and Table S6, referral provision exhibits assortative patterns. In particular, referrals are more common among people in the same age range, whereas nepotism often involves individuals from different age groups (children of relatives) (Wang, 2013; Foley, 2014).

4 Referral Benefits To Workers and Firms

4.1 Referral Benefits to Workers

This section examines whether referrals improve referees' labor market outcomes, conditional on finding a job. Our framework for analyzing the benefit of referrals is conceptually similar

[24]We examine the preference argument in more detail in Section 5.

to model (2):

$$Y_{ilr} = \beta \text{Friend}_{ilr} + \boldsymbol{X_i Z_l \gamma} + \lambda_c + \alpha_r + \varepsilon_{ilr}, \tag{3}$$

where Y_{ilr} denotes the labor market outcome of worker i who switches to work location l in neighborhood c and lives in residential neighborhood r. We control for the same set of demographic variables and location attributes considered in model (2). Because we do not observe individuals' socioeconomic background and status, such as education and wealth, we include in all regressions the residential neighborhood fixed effect (α_r) as a proxy (luxurious complexes versus low-income neighborhoods).

We construct five different measures of job quality. The first one is the expected wage at the new job, measured by the average annual payroll (in thousand RMB) among firms in the same location weighted by their number of employees. Wage dispersion is often driven by across-firm rather than within-firm variations (Card et al., 2018). In addition, an individual's housing value is correlated with his labor income. Thus, we use *coworkers'* housing price as a second measure for job compensation. Specifically, we construct the difference between the average housing price of coworkers at the new workplace and that of coworkers at the previous job. Large positive differences are more likely to be associated with increases in wages and other pecuniary benefits.

The other three measures of job amenities include whether the move is from a part-time to a full-time job, reductions in commuting distance, and whether the move is from a non-SOE (state-owned enterprise) firm to an SOE, because openings at SOEs are sought after for their job security and pension benefits (Zhu, 2013).[25] Although none of these measures of job outcomes is perfect, collectively they speak to both the financial and nonfinancial aspects of job quality.

Results Since our labor market outcomes are constructed from different data sources, the number of observations across specifications in Table 6 varies from 15,881 to 29,117 and reflects the varying extent of missing information. Referral jobs pay higher expected wages than nonreferral jobs. The point estimate of the wage premium is RMB 620, or about 2 precent of the average wage reported in our sample.[26] Turning to differences in coworkers' home values in the new versus old workplace, referral jobs are associated with a 0.5 percent

[25]State owned enterprises (SOEs) account for a small fraction of the total number of firms, but they constitute more than 30 percent of China's GDP and 20 percent of total employment (State Assets Supervision and Administration Commission 2017). Many SOEs appear in the Fortune Global 500 list and are among the largest conglomerates in the world. Private and foreign companies trail behind SOEs in terms of firm size and revenue. Employment opportunities at SOEs are coveted for their job security, generous benefits, and sometimes higher wages than those in nonstate sectors. A workplace is classified as "SOE" if the majority of workers at that location are employed by SOE firms.

[26]The annual wage is measured in thousand RMB and the mean is 31.

higher housing price per square meter (the average housing price in the city is RMB 13,000 ($2,000) per square meter.)

Having at least one friend at the new workplace increases the probability of moving from a part-time to a full-time job by 1.4 percentage points, which is a 2 percent increase in the likelihood of working full time.[27] 31 percent of job changes involve a shorter commute. Referred jobs are associated with a 30 percent increase in the likelihood of working closer to home. Finally, having a friend at an SOE firm raises the probability of moving there by 1.2 percentage points, an 11 percent increase from the mean (0.11). Higher wages are an indication of enhanced worker productivity, and shorter commutes and full-time positions reflect better job amenities. In Appendix Table S7, we repeat our analysis using an alternative definition of friends that includes all social contacts with at least four weeks of nonmissing work locations. Referral jobs are associated with a 1.3 percent increase in wage premium, a 0.6 percent increase in job-related benefits (as proxied by coworkers' housing prices), and a 12 percent increase in the likelihood of working full time, similar to that found in the baseline specification. The effects on the likelihood of a shorter commute and transitioning to an SOE firm are also very similar.

Our results provide evidence that referrals lead to better matches between workers and vacancies. These patterns are also consistent with the hypothesis that referrals improve workers' labor market outcomes through mitigating information frictions in the hiring process.

4.2 Referral Benefits to Firms

With a few exceptions, most empirical studies on job referrals abstract away from analyzing firm outcomes, because comprehensive data on the performance of both employees and employers are hard to obtain.[28] We merge the call data with administrative firm-level data based on locations and examine variation across a large number of firms in different industries.

We successfully merge between 5,000 and 10,000 firms, 67 percent of which are manufacturing firms that require production facilities.[29] Our main specification focuses on locations matched to large firms that have more than 100 employees, which represent about 20 percent

[27] Hours worked is derived from phone usage during workdays at the workplace and is conservative by the nature of such records. Part time (full time) is defined as 30 hours or less (more than 30 hours). On average, 57 percent of the switchers work full-time before the job change, reflecting the conservativeness of this measure.

[28] A notable exception is Burks et al. (2015) that use data from nine large firms in three industries (call centers, trucking, and high-tech) to analyze whether firms benefit from referrals.

[29] The exact number of successful merges is withheld to keep the city anonymous.

of our sample. The average employment for these large firms is 150; thus, these firms are likely to occupy an entire location. While limiting to large firms significantly reduces the sample size, it reduces the likelihood of erroneously linking workers to unrelated firms.[30] Table S8 reports results from replicating the analysis using all firms. Results are similar both statistically and economically, which is reassuring. In the rest of this section, we use "location" and "firm" interchangeably.

We compare the performance of firms that hire through referrals to firms that hire through other channels via the following model:

$$Y_i = \gamma \text{Referral}_i + \boldsymbol{Z_i \beta} + \lambda_c + \varepsilon_i \tag{4}$$

where i denotes a firm. We examine three measures of firm performance Y_i: (1) net inflow of workers, or the number of hires minus separations; (2) match rate, measured by net inflow over vacancies; and (3) firm growth rate, measured by net inflow over total number of employees.[31] We limit our analysis to locations with at least one hiring, otherwise the estimate of γ will be artificially inflated since the number of hires is at least one for locations with referrals by construction.

Referral_i is a dummy variable that takes value one if at least one worker who switches to firm i has a friend working there, while λ_c denotes neighborhood fixed effects – the same as in model (3). $\boldsymbol{Z_i}$ denotes firm attributes and employee characteristics. Firm attributes include age, dummies for 18 different industries, a dummy for SOEs, the average number of employees (firm size) and real capital from 2010 to 2015. To capture preexisting trends, we also control for the average employment growth rate from 2010 to 2015. In addition, we include a firm's referral network size, defined as the number of unique social contacts owned by employees who work in firm i prior to the arrival of new hires. Worker attributes include the shares of female workers and migrants, the average age of employees, and the average housing price of preexisting employees.

Results The parameter estimate γ captures the effect of using referrals on firms' performance. The dependent variables in Table 7 are in logs, hence γ is directly semi-elasticities: the percentage change in the outcome variable when firms hire through referrals. To the extent that firms that grow quickly are more likely to hire through employee referrals, our estimate could be biased upward. To tackle this problem, we estimate model (4) with an

[30] For the same reason, we repeated the referral analysis limiting to locations matched with large firms (that have more than 100 employees). The results are very similar to the baseline estimates.

[31] There are two measures of worker inflows: gross inflows and net inflows. Results reported below use net inflows, though they are similar to gross inflows.

increasingly rich set of variables that control for firm growth and employee quality.

The Referral$_i$ coefficient estimates are remarkably similar across different sets of controls for firm and employee attributes. Firms that recruit through referrals are associated with more hires, better matching rates, and higher growth rates: using referrals increases a firm's net labor inflow by 63 percent, enhances the job matching rate by 86 percent (the average matching rate for large firms is 0.76), and raises the firm growth rate by 45 percent (the median growth rate is 4 percent for large firms).[32] Results in Table S8 that use all firms document similar patterns. We replicated our analysis with various other selection criteria (for example, using all friends with at least four weeks of nonmissing work locations as in Table S9, as well as friends with three months or six months of work locations) and obtained very robust findings.

Although our analysis in this section is descriptive because we lack suitable instruments, the estimates' robustness to a rich set of firm and worker controls raises our confidence that these estimates are not simply picking up unobserved firm and employee quality. Instead, firms are likely to benefit from employee-provided referrals, consistent with the fact that referral-based hiring programs are common (Burks et al., 2015).

5 Robustness Analysis of the Referral Effect

This section conducts an extensive set of robustness checks for the referral effect estimated above. We first evaluate the importance of unobserved location attributes, followed by homophily and preferences for working with friends, before examining reverse causality concerns and robustness to different friend definitions.

Local Demand So far our regressions have focused on individuals' and their social ties' characteristics and location amenities (such as number of restaurants and parking lots). One might be concerned about local confounders such as differences in local labor market demand and unobserved location attributes that are correlated with the referral dummy.

Table 8 repeats the referral analysis with an increasingly rich set of location attributes. In addition to the regressors in the baseline specification, Column 1 controls for the number of vacancies in each occupation within a 500-meter radius of a location and their interactions with switcher i's demographics (gender, age group, and migrant status). These interactions capture potentially heterogeneous labor demand. Column 2 further includes the number of employees and firm attributes in each location, such as firm age, revenue, real capital stock,

[32] The matching rate can exceed one as not all vacancies are posted (the data coverage on vacancies is incomplete).

average payroll per employee, as well as the interaction of these variables with switcher i's demographics. In light of the evidence in Section 2.2 on the strong parallel movement between information flows and worker flows, Column 3 adds the total number of calls made or received by individuals working in location l, excluding switcher i. This captures unobserved location attributes (for example, proximity to commercial centers) that may affect job changes. Column 4 further incorporates the call volume among people living in location l.

The last robustness check on unobserved location attributes relates to our falsification test in Table 4, where the coefficient for friends who moved away is much smaller (0.07) than that for referral friends (0.35). We interpret the coefficient difference as evidence that referrals provide job information, but an alternative explanation could be that friends move away because of an undesirable (and time-varying) work environment that reduces switchers' probability of moving to that location. To examine this directly, Column 5 further controls for turnover at the new work location prior to the job switch. Specifically, we include as an additional regressor the ratio of the average monthly number of people leaving location l during the three months prior to the job change over the average monthly turnovers at location l for the rest of the sample period.

Reassuringly, both the referral estimate and the coefficient for "friends moving away" are stable across different columns of Table 8, indicating that the baseline estimate is not driven by unobserved location attributes or differences in local labor market features.

Homophily Table 4 provides evidence against homophily by comparing different types of friends (for example, friends who recently moved away and friends living but not working at location l). To examine homophily in further detail, Table 9 directly controls for observable attributes that reveal similarity between switchers and referrals. This exercise is in spirit similar to the literature that uses observables to gauge the nature of unobservables (Altonji et al., 2005; Oster, 2019). If switchers are similar to their referral friends in observable attributes, then they are likely to be similar in dimensions that we do not observe (notably, work attitude and preference). The richness of our data allows us to use not only characteristics of switchers and referrals but also characteristics of their social networks. Our test is inspired by the literature on social network formation (Fafchamps and Gubert, 2007) showing that homophily (common preferences, tastes and attitudes) is the main driver of friendship formation. It follows that individuals sharing larger number of common friends and friends with common features should have higher levels of homophily than individuals sharing a lower overlap. Hence including as regressors both individual characteristics and their social ties' attributes should be a direct and effective way to control for homophily.

Column 1 controls whether the referral at location l has the same gender as switcher i, is in the same age group and from the same birth county, and has similar wealth (proxied by housing price).[33] As "birds of a feather flock together," Columns 2 and 3 further control for the similarity in switchers' and their referrals' social network. Column 2 adds the share of mutual friends and the share of common work neighborhoods that are covered by i's and l's social network.[34] Column 3 first calculates the fraction of social ties who are female, migrant, in each age group and birth county, and their average house price for both switchers and their referrals. Then it controls for differences in these demographics (and wealth) between switchers' social ties and referrals' social ties, in addition to the controls in Columns 1 and 2. If switcher i and referral l mingle with similar friends, then they are likely to be similar as well.

Instead of using parametric functions of observables as in Columns 1–3, Columns 4–6 use a popular unsupervised machine learning algorithm, the K-means clustering algorithm, to nonparametrically profile switchers and their social ties (including referral friends).[35] The K-means clustering is performed based on an individual's own attributes (gender, age, whether born in the local city or local province, house price), as well as his social ties' attributes (number of neighborhoods covered by his friends, share of friends who are female, migrants, in each age bin, and their average house price). We group switchers and their social ties into 5, 10, and 100 clusters, respectively, and control for a dummy variable indicating whether the referral friend at location l is in the same cluster as switcher i.

Across all columns of Table 9, the estimated referral effects are stable and vary between 0.32 and 0.34. While they are slightly lower than the baseline estimate of 0.35, the magnitude is quite reasonable and indicates that the baseline controls (which include a rich set of demographic and location-attribute interactions and old-by-new neighborhood-pair fixed effects) are adequate at capturing sorting and that our results are not driven by homophily.

Preferences for Working with Friends Another potential explanation for the estimated referral effect is preference for working with friends. For example, workers might prefer

[33] The house price is "similar" if the difference is within one standard deviation of the housing price distribution. In locations with more than one friend, friend attributes are constructed using the average.

[34] The share of mutual friends and the share of common work neighborhoods are calculated using the Jaccard index: $J(A,B) = \frac{A \cap B}{A \cup B}$. For example, the share of mutual friends between i and l is equal to the number of mutual friends divided by the total number of unique friends among i's and l's social ties.

[35] K-means clustering uses iterative procedures to partition data into k nonoverlapping groups or clusters. The procedure begins with k randomly picked initial group centers. Each individual is assigned to the group with the closest center to minimize the within-group Minkowski distance metric with argument 2 (i.e., the L2 Euclidean distance) along characteristics. The mean of the observations assigned to each of the groups is computed, and the process is repeated until all observations remain in the same group from the previous iteration.

to co-locate with their friends (even if there is no job-related information shared among them). Park (2019) uses a field experiment and shows that people are willing to forego 6 percent of wage to work with friends. While preference is a plausible and potentially relevant explanation, individual preferences are rarely explicitly measured and difficult to examine using observational datasets.

It is worth noting that the robustness analyses on homophily discussed above are informative about the preference argument, since homophily – being attracted to others who are similar to oneself – is closely related to preference. Our referral estimate survives a extensive set of controls on switchers' and referrals' attributes and the attributes of their social ties, indicating that the referral effect we estimate is not driven by people's preference to work with friends who are similar to them.

To further examine the importance of preferences for working with friends, we leverage the spatial variation in switchers' social networks.[36] If people have strong preferences to be with friends, then we should expect that switchers are more likely to move to places with more friends, everything else equal. Column 1 of Table 10 follows the baseline specification but controls for the number of friends in addition to the friend dummy. In the second analysis, we compare the number of friends in the new workplace with that in the old workplace. To see why this is relevant, consider two individuals, switchers A and B, both of whom have friends in their new workplace. Switcher A has fewer friends in his old workplace than B. If the preference to mingle with friends is important, then we might expect switcher A to exhibit a stronger preference for the new work location. Column 2 of Table 10 controls for a dummy that takes value one if the new workplace contains more friends than switcher i's old workplace. Columns 3 and 4 replicate Columns 1 and 2 but also control for friend attributes in the new workplace (Column 3) and friend attributes in both the new and old workplace (Column 4).

The coefficient on the number of friends indicates that indeed people prefer to mingle with friends. However, the coefficient is small in magnitude (and insignificant): having one more friend increases the switching probability by 0.005, much smaller than the estimated referral effect that increases the switching probability by 0.34. Similarly, having more friends in the new workplace than at the old workplace modestly increases the switching probability by 0.001 to 0.002. In both cases, the referral effect estimate is stable and robust to these additional controls. These results suggest that the preference to work with friends is not a major consideration in our setting.

[36] Among all location-switcher pairs that contain at least one friend, 69 percent of locations have exactly one friend, 16 percent have two friends, and 15 precent have three or more friends.

Reverse Causality Our analysis defines switcher i's social network as the one that is formed three months prior to his job switch. As discussed in Section 2.2, the three-month cutoff is chosen to reflect the typical length of job searches. It also mitigates the concern of a reverse causality where some social ties in location l are established after switcher i has found a job there. Table S10 uses the baseline specification but with an increasingly stringent cutoff to define switchers' social network – from one month prior to the job change (Column 1) to five months prior to the job change (Column 5). The referral effect is robust to different cut-offs, consistent with the fact that few links are formed immediately before the job switch.

Alternative Friend Definition We conducted a few additional robustness checks using different friend definitions (Table S11). Our baseline analysis limits to friends who have at least 45 weeks of nonmissing work locations. This mitigates measurement errors in friends' job locations, but omits a large fraction of friends for whom we observe fewer than 45 weeks of location information. The first column of Table S11 replicates the baseline analysis (Column 2 of Table 4) using all friends who have at least four weeks of nonmissing work locations. This enlarges the number of individual-friend pairs from 401,437 to 979,595. The estimated referral effect remains robust: having a friend in a location increases the probability of moving there by 36 percentage points.

Social ties are one-way contacts in the baseline analysis. Column 2 of Table S11 defines individual i's friends as social contacts with two-way communications: people who both make phone calls to and receive phone calls from individual i. In addition, all friends with at least four weeks of nonmissing work locations are included in the analysis. The referral estimate is slightly higher than that of our base specification (0.38). As work locations are missing for friends outside Company A's subscriber network, one might be concerned about potential sample selection biases. Columns 3 and 4 split the switcher sample based on whether the friend coverage is above or below the median (the cut-off is 48 percent). The difference in the referral estimates between these two subsamples is modest and insignificant.

6 Comparison with Literature and External Validity

We conclude our analysis with some extensions that shed light on the external validity of our findings. We first compare results with the existing literature and validate proxies for social interactions that are commonly used in the literature. Then we examine how call volumes relate with other communication channels such as text messages and mobile apps (WeChat). Finally, we repeat the analysis for individuals who experienced unemployment spells but then subsequently found a new job within our sample period.

Proxies for Social Ties How do our results compare to referral measures in the existing literature? Two common approaches of inferring social networks are used in observational studies. The first, pioneered by Bayer et al. (2008), defines referrals as residential neighbors. Using data from the Boston metropolitan area, they treat as friends individuals who live in the same census block and examine the cross-sectional correlation between place of work and place of residence – whether friends are more likely to work in the same census block than individuals living in the same census block group or 10 nearest blocks but not in the same block. The second approach assumes that social interactions are stronger within an ethnic group and defines friends as coworkers who are members of the same minority group (Bandiera et al., 2009; Dustmann et al., 2016). We re-estimate model (2) using these alternative definitions of friendship and report the results in Table 11. "Residential neighbor" is a dummy variable that takes value one if workplace l contains at least one individual who shares the same residential location as i. Ethnicity, which is inapplicable in China's context, is replaced with birth county as the literature documents strong social ties among individuals from the same birth region (Zhao, 2003). "Same birth county" takes value one if individual i has a coworker in location l who was born in the same county. Columns 1 and 2 only include these alternative definitions of friends. Column 3 contrasts neighbors with referral friends who are not neighbors, while Column 4 compares coworkers who share the same birth county with referral friends who work in the same location but have different birth counties.

The results in Table 11 confirm the findings in the literature that neighbors and coworkers from the same birth counties are important. The coefficients on neighbors and the same birth county are 0.21 and 0.10, respectively, when they are the only measure of an individual's social network. Given the average moving probability of 0.09, having a social tie of either type significantly increases the probability of switching to location l. On the other hand, the effect of our friend measure dominates both types of social ties by a large margin. The difference in coefficient magnitude is both statistically significant and economically sizable, and in the case of "same birth county", the effect of our friend measure is four times as large (Column 4).

To further validate these different referral proxies, we use the disaggregated call records to directly examine the communication intensity between neighbors, coworkers, and individuals sharing the same birth county. In Column 1 of Table S12, we randomly draw 1 percent of individuals (including both switchers and nonswitchers) and examine the average monthly calls between pairs of individuals living in the same residential neighborhood and pairs of individuals living in the same location within a neighborhood. This mimics the empirical setting in Bayer et al. (2008) that contrasts individuals living in the same census block with individuals living in the same census block-group but not the same block. Residential neighborhood fixed effects are included to control for observed and unobserved neighborhood characteristics. Conditional on living in the same residential neighborhood, neighbors living in the same location make 4.5 times as many calls as two random individuals residing in the same neighborhood. In Column 2, we compare call frequency between coworkers versus call frequency between pairs of individuals working in the same neighborhood. Similar to the finding on neighbors, coworkers on average make 4 times as many phone calls as two random individuals working in the same neighborhood. Column 3 uses the same sample as Column 2 but examines the importance of sharing the same birth county. People with the same birth county and working in the same neighborhood do communicate more frequently than individuals born in different counties, though the effect size is much smaller than the other two types of social ties.

In sum, our results support the analysis of social interactions using neighbors, coworkers, and same-ethnicity individuals as proxies for social ties, but suggest that the estimated referral effects using these proxies are likely to be a lower bound.

Weak vesus Strong Ties The literature on weak versus strong ties in general finds that the referral effect from a strong tie is stronger than that from a weak tie (Gee et al., 2017b; Bian, 1997). We measure the tie strength by call intensity and revisit this question in Table S13. We follow the baseline specification (Column 2 of Table 4) and include interactions between the "Friend" dummy and measures of call intensity. In Column 1, "Call intensity" is the demeaned number of calls between switcher i and his referral friend at location l prior to the job switch. In Column 2, "$Call_{il}/Call_i$" is a demeaned ratio of the number of calls between switcher i and location l as a fraction of all calls made by i prior to the job switch. This call-frequency ratio takes into consideration differences across individuals (some people are more outgoing than others) and is a better measure of tie strength. The coefficient of the interaction term is positive and significant in both columns, suggesting that the referral effect strengthens with tie strength. For example, a one standard deviation increase in $Call_{il}/Call_i$ is associated with a 6 percent increase in the referral effect.[37]

Online Appendix Figure S5 illustrates graphically the relationship between tie strength (measured by $Call_{il}/Call_i$) and the referral effect. Following Gee et al. (2017b), we limit to switchers who found a job through a referral friend (or referees). The top panel plots

[37] To calculate this, note that the s.d. of $Call_{il}/Call_i$ is 0.06. While the referral effect increases with the strength, it is significant even among weak social ties, as reflected by the large friend coefficient. Figure S6 in the Online Appendix presents the event study controlling for tie strength. The communication pattern between referral pairs remains hump-shaped.

the distribution of tie strength between referees and all of their social ties. The middle panel plots the distribution only for referee-referral pairs. Most social ties are weak ties: tie strength is lower than 0.05 for 90 percent of social ties and 75 percent of referral ties. The bottom panel plots the predicted probability that switcher i moves to friend l's workplace using a linear regression $P_{il} = c + b * \text{TieStrength}_{il} + \epsilon_{il}$. Corroborating Table S13, the stronger the relationship with a friend, the larger the referral effect. These patterns are consistent with the evidence in the existing literature, despite the differences in tie strength measures and data contexts.

Phone Calls and Other Communication Channels A key premise of our analysis is that call volume serves as a proxy for the amount of information exchanged among individuals. In practice, there are many different communication channels, such as text messages, emails, and apps (WeChat), in addition to phone calls. One potential concern with our analysis is that people might be using text messages or WeChat in lieu of phone calls to communicate with friends. If there is a negative correlation between phone calls and alternative information channels, then observing a high phone call volume does not imply more information exchange between individual i and his social contacts, as the increased call volume could be offset by reduced text messages and app usage.

The most important alternative communication channel in our context is WeChat, the dominant social media app in China with 1.2 billion users as of 2020 (Tecent, 2020). We do not observe WeChat usage for individuals in our sample and thus consider three alternative measures. We repeat the analysis for all cellphone users as well as job switchers separately. Since 3G service is too slow for WeChat, we use whether an individual's phone is compatible with the 4G network as a proxy for WeChat usage.[38] On average, individuals with 4G devices make 67 more calls per month than individuals with 3G devices. Job switchers using 4G-compatible devices make 59 more calls. Regressing an individual's average monthly phone calls on his phone device's 4G compatibility leads to a positive, significant, and economically large coefficient for both switchers and nonswitchers. The pattern is robust across all specifications we examined with an increasing set of controls for neighborhood fixed effects, individual demographics, phone attributes and mobile plans (Panel A of Table 12). Similar to the findings with the 4G compatibility, there is a positive and significant correlation between monthly calls and the allowed Internet data volume in one's cellphone plan (Panel B). Finally, we observe individuals' internet browsing behavior (duration in thousand minutes) during the second week of May 2017. Consistent with the evidence above, individuals' number of calls is positively correlated with the time they spent surfing the Internet in the same

[38]Two-thirds of mobile phones in our sample are compatible with the 4G network.

week (Panel C). These patterns hold for both regular cell phone users and job switchers in particular.

To directly examine individual usage of text messages and WeChat, we collect additional information from 20,000 randomly selected cellphone users in a comparable city in China on WeChat usage for four weeks from October to November 2020 and on the number of text messages for one week in November 2020. Our findings are presented in Table S14. Individuals who talk more (measured by either the number of calls or call duration) also send more text messages and use WeChat more frequently. More importantly, the positive correlation also survives individual fixed effects: a user talks more in weeks when he utilizes WeChat more frequently.

These patterns suggest that different communication channels are complements: individuals who make more phone calls also send more text messages, use WeChat more intensively, and browse the Internet more. As a result, while our call records do not cover other information channels that are available to an individual, they serve as a good proxy for the amount of information exchanged between an individual and his/her social ties.

To examine the referral effect's robustness to different information channels, we repeat our baseline analysis separately for switchers whose phones are compatible or incompatible with the 4G network (Table S15). While the parameter estimate is marginally higher for 4G users who are much more likely to use WeChat than 3G users, the difference (smaller than 0.01) is statistically insignificant. Thus, referrals are at work for both groups of individuals, whether or not they can use WeChat.

Switchers with Employment Gaps Our data also provide an opportunity to examine the referral effect for switchers with employment gaps. However, it is considerably more challenging to distinguish unemployment from other factors that also lead to intermittent work location patterns (travel, sick leave, part-time jobs, etc.). In addition, the referral analysis for switchers with employment gaps is limited to employed individuals who experience unemployment and then manage to find another job within our sample period (and hence experience a relatively short unemployment duration). In the end, we identify a total of 3,638 individuals with one employment gap, of whom 1,677 find jobs through referrals (see Online Appendix Section S4 for more details). The modest sample size reflects challenges in measuring unemployment; as a result, we interpret the results below using reemployed individuals as suggestive. Nonetheless, this analysis could be informative about the generalizability of our main results.

We begin with an event study to examine the communication patterns between reemployed individuals and their referrals at different stages (before being unemployment, during unemployment, and post reemployment) and contrast with the communication patterns between these reemployed individuals and their nonreferral friends:

$$\begin{aligned}\text{Freq}_{ijt} = \quad & \textstyle\sum_{s=-5}^{4} \gamma_s \text{Referral}_{ij} \cdot 1[t = s, r < 0] + \sum_{r=0}^{4} \gamma_r \text{Referral}_{ij} \cdot 1[t = r, s \geq 0] + \\ & \textstyle\sum_{s=-5, s\neq -1}^{4} b_s \text{nonreferral}_{ij} \cdot 1[t = s, r < 0] + \sum_{r=0}^{4} b_r \text{nonreferral}_{ij} \cdot 1[t = r, s \geq 0] \\ & + \lambda_i + \tau_t + \epsilon_{ijt}\end{aligned}$$

As there are two events (the unemployment and the reemployment), we use s to denote the event window index for unemployment and r to denote the index for reemployment. Similar to the analysis using on-the-job switchers, the reference category is the call frequency between reemployed individuals and their nonreferral friends during the month immediately before the unemployment, $s = -1$. The period prior to the fifth month before unemployment is binned with month -5, the period after the fourth month post unemployment is binned with month 4, and the period after the fourth month post reemployment is binned with month 4. The event study is presented in Figure 4, which exhibits similar patterns as Figure 3. Most notably, the communication pattern with referrals has an inverted-U shape during unemployment (the period of job search) that peaks prior to the reemployment. Similar to on-the-job switchers, people with employment gaps also experience more intense communication with their referrals before finding a new job. We have repeated the event study using different unemployment definitions. The qualitative pattern of more pronounced communication between people searching for jobs and their referrals during the search period (the unemployment spell) is present in all event studies we conducted (Figure S7).

Similar to on-the-job switchers, people who experience employment gaps also benefit from referrals (Table S16). The referral effect for this population varies from 0.31 to 0.33, in line with our findings in Section 3.

In Table S17, we examine whether switchers during unemployment spells search more actively for job opportunities than on-the-job switchers. We define the search period as the unemployment spell for individuals with employment gaps and the three months prior to the job change for on-the-job switchers. The regression sample consists of all switchers with and without unemployment spells. Column 1 regresses average monthly calls during the search period on a dummy that takes value one for people who experience unemployment spells. Individuals during unemployment make on average 40 more calls per month (with a sample mean of 133 calls per month) than on-the-job switchers. Columns 2–6 control for the talking frequency during nonsearch periods. The R-squared increases from 0.01 in Column 1 to 0.50 in Column 2, indicating the importance of personality in explaining variations in call volume. Nonetheless, the unemployed coefficient only changes modestly from 40 to 36 calls. Columns 3 and 4 further add demographic controls and residential neighborhood fixed effects. The estimate remains stable: unemployed individuals make on average 33–34 more calls. Columns 5 and 6 split the sample into two, based on whether the unemployment spell is below or above the sample median (11 weeks). People experiencing longer unemployment spells talk slightly more, though the difference is statistically insignificant. To the extent that call frequency reflects search intensity, these communication patterns provide suggestive evidence that individuals exert more effort searching for jobs when they are unemployed.

7 Conclusion

This paper provides the first direct evidence of increased communication around job changes among referral pairs. We use geocoded mobile phone records to illustrate that information provided by social contacts mitigates information asymmetry and improves labor market performance.

Our study provides three broad lessons for future research. First, panel data with fine spatial and temporal variation hold great potential for overcoming the challenges of causal inference with observational data in the context of social networks. For example, the ability to identify different types of social contacts in small geographical areas at overlapping periods helps us to tackle sorting. Second, big data from nonconventional sources complement well traditional datasets on socioeconomic measures. In our analysis, tax records and firm registration data are crucial to study how referrals benefit firms, a topic that is understudied in the existing literature. Third, information exchange, and in particular, social and socioeconomic diversity in communication appears to facilitate worker movement. In the future, studies on the exact mechanism that governs how information exchange through referrals increases labor market efficiency would be extremely valuable.

References

Abramitzky, R. and Boustan, L. (2017). Immigration in american economic history. *Journal of Economic Literature*, 55(4):1311–45.

Akbar, P. A., Couture, V., Duranton, G., and Storeygard, A. (2018). Mobility and congestion in urban india. Working Paper 25218, National Bureau of Economic Research.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism*. *The Quarterly Journal of Economics*, 84(3):488–500.

Alesina, A., Harnoss, J., and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21(2):101–138.

Altonji, J., Elder, T., and Taber, C. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184.

Aral, S. and Alstyne, M. V. (2011). The diversity-bandwidth trade-off. *American Journal of Sociology*, 117(1):90–171.

Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290.

Ashraf, Q. and Galor, O. (2011). Cultural diversity, geographical isolation, and the origin of the wealth of nations. Working Paper 17640, NBER.

Aslund, O., Hensvik, L., and Skans, O. (2014). Seeking similarity: How immigrants and natives manage in the labor market. *Journal of Labor Economics*, 32(3):405–41.

Athey, S., Blei, D., Donnelly, R., Ruiz, F., and Schmidt, T. (2018). Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *AEA Papers and Proceedings*, 108:64–67.

Bai, J. (2018). Melons as lemons: Asymmetric information, consumer learning and quality provision. Working paper.

Bailey, M., Cao, R., Kuchler, T., and Stroebel, J. (2018a). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018b). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80.

Baker, S. R. (2018). Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy*, 126(4):1504–1557.

Bandiera, O., Barankay, I., and Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094.

Barahona, N., Otero, C., Otero, S., and Kim, J. (2021). Equilibrium effects of food labeling policies. *Stanford Working Paper*.

Bayer, P., Ross, L. S., and Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–96.

Beaman, L. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the u.s. *Review of Economic Studies*, 79(1):128–61.

Bian, Y. (1997). Bringing strong ties back in: Indirect ties, network bridges, and job searches in china. *American Sociological Review*, 62(3):366–385.

Bjorkegren, D. and Grissen, D. (2018). Behavior revealed in mobile phone usage predicts credit repayment. *World Bank Economic Review*. Accepted.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Blumenstock, J., Chi, G., and Tan, X. (2019). Migration and the Value of Social Networks. CEPR Discussion Papers 13611, C.E.P.R. Discussion Papers.

Blumenstock, J. E., Cadamuro, G., and On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350:1073–1076.

Buchel, K., Ehrlich, M. V., Puga, D., and Viladecans-Marsal, E. (2019). Calling from the outside: The role of networks in residential mobility. Working Paper wp2019-1909, CEMFI.

Burks, S. V., Cowgill, B., Hoffman, M., and Housman, M. (2015). The value of hiring through employee referrals. *Quarterly Journal of Economics*, pages 805–839.

Cai, H., Fang, H., and Xu, L. C. (2011). Eat, drink, firms, government: An investigation of corruption from the entertainment and travel costs of chinese firms. *The Journal of Law and Economics*, 54(1):55–78.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70.

Cingano, F. and Rosolia, A. (2012). People i know: Job search and social networks. *Journal of Labor Economics*, 30(2):291–323.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley–Interscience.

Dai, R., Mookherjee, D., Munshi, K., and Zhang, X. (2018). Community networks and the growth of private enterprise in china. *VoxDev*.

Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4):171–98.

Dustmann, C., Glitz, A., Schonberg, U., and Brucker, H. (2016). Referral-based Job Search Networks. *The Review of Economic Studies*, 83(2):514–546.

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

Edin, P.-A., Fredriksson, P., and Aslund, O. (2003). Ethnic enclaves and the economic success of immigrants: Evidence from a natural experiment. *The Quarterly Journal of Economics*, 118(1):329–357.

Evtushenko, A. and Kleinberg, J. (2021). The paradox of second-order homophily in networks. *Scientific Reports*, 11(1):13360.

Fafchamps, M. and Gubert, F. (2007). Risk sharing and network formation. *American Economic Review*, 97(2):75–79.

Foley, M. (2014). *JPMorgan Gives Regulators Evidence of Nepotism in China*. Retrieved March 13, 2014, from https://www.cheatsheet.com/money-career.

Gagnon, J., Xenogiani, T., and Xing, C. (2014). Are migrants discriminated against in chinese urban labour markets? *IZA Journal of Labor & Development*, 3(1):17.

Gee, L., Jones, J., and Burke, M. (2017a). Social networks and labor markets: How strong ties relate to job finding on facebook's social network. *Journal of Labor Economics*, 35(2):485–518.

Gee, L. K., Jones, J. J., Fariss, C. J., Burke, M., and Fowler, J. H. (2017b). The paradox of weak ties in 55 countries. *Journal of Economic Behavior & Organization*, 133:362–372.

Giltz, A. (2017). Coworker networks in the labor market. *Labour Economics*, 44:218–230.

Glaeser, E. L., Kominers, S. D., Luca, M., and Naik, N. (2015). Big data and big cities: The promises and limitations of improved measures of urban life. Working Paper 21778, National Bureau of Economic Research.

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453:779.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380.

Hastings, J. S. and Weinstein, J. M. (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *The Quarterly Journal of Economics*, 123(4):1373–1414.

Hellerstein, J., McInerney, M., and Neumark, D. (2011). Neighbors and coworkers: The importance of residential labor market networks. *Journal of Labor Economics*, 29(4):659–95.

Hellerstein, J., McInerney, M., and Neumark, D. (2014). Do labor market networks have an important spatial dimension? *Journal of Urban Economics*, 79(4):39–58.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2):994–1028.

Hoffman, M. (2017). The value of hiring through employee referrals in developed countries. *IZA World of Labor*.

Ioannides, Y. M. and Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42:1056–1093.

Jeffers, J. (2018). The impact of restricting labor mobility on corporate investment and entrepreneurship. Working paper, SSRN.

Jin, G. Z. and Leslie, P. (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, 118(2):409–451.

Kramarz, F. and Skans, O. N. (2014). When Strong Ties are Strong: Networks and Youth Labour Market Entry. *The Review of Economic Studies*, 81(3):1164–1200.

Kreindler, G. E. and Miyauchi, Y. (2019). Measuring commuting and economic activity inside cities with cell phone records. Working paper.

Munshi, K. and Rosenzweig, M. (2013). Networks, commitment, and competence: Caste in indian local politics. Working Paper 19197, National Bureau of Economic Research.

National Bureau of Statistics of China (2014). *The Third National Economic Census Report*. Retrieved December, 2014, from http://www.stats.gov.cn.

Ngai, L. R., Pissarides, C. A., and Wang, J. (2017). China's mobility barriers and employment allocations. *Journal of European Economic Association*. Forthcoming.

Nie, P. and Sousa-Poza, A. (August 2017). What chinese worker value: An analysis of job satisfaction, job expectations and labor turnover in china. *IZA Discussion Papers No. 10963*, 108.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Ottaviano, G. I. and Peri, G. (2006). The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography*, 6(1):9–44.

Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. Princeton University Press.

Park, S. (2019). Socializing at work: Evidence from a field experiment with manufacturing workers. *American Economic Journal: Applied Economics*, 11(3):424–55.

Recchi, E. (2009). Chapter 4: The social mobility of mobile europeans. volume 1 of *Pioneers of European integration citizenship and mobility in the EU: Citizenship and mobility in the EU*, pages 72 – 97. Edward Elgar Publishing.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649.

Saygin, P. O., Weber, A., and Weynandt, M. A. (2018). Coworkers, networks and job search outcomes. *ILR Review*. Forthcoming.

Schmidheiny, K. and Siegloch, S. (2020). On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *ZEW - Centre for European Economic Research Discussion Paper No. 20-017*.

Schmutte, I. (2015). Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics*, 33(1):1–33.

Schmutte, I. (2016). How do social networks affect labor markets? *IZA World of Labor*.

Smith, V. K. and Johnson, F. R. (1988). How do risk perceptions respond to information? the case of radon. *The Review of Economics and Statistics*, 70(1):1–8.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Tecent (2020). *Tecent Announces 2020 Third Quarter Results*. Retrieved November, 2020, from https://www.tencent.com/en-us/investors/financial-news.html.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies*, 68(2):261–295.

Topa, G. (2011). Labor markets and referrals. *Handbook of Social Economics*, 1B:1193–1221.

Wang, S. Y. (2013). Marriage networks, nepotism, and labor market outcomes in china. *American Economic Journal: Applied Economics*, 5(3):91–112.

Whalley, J. and Zhang, S. (2007). A numerical simulation analysis of (hukou) labour mobility restrictions in china. 83(2):392–410.

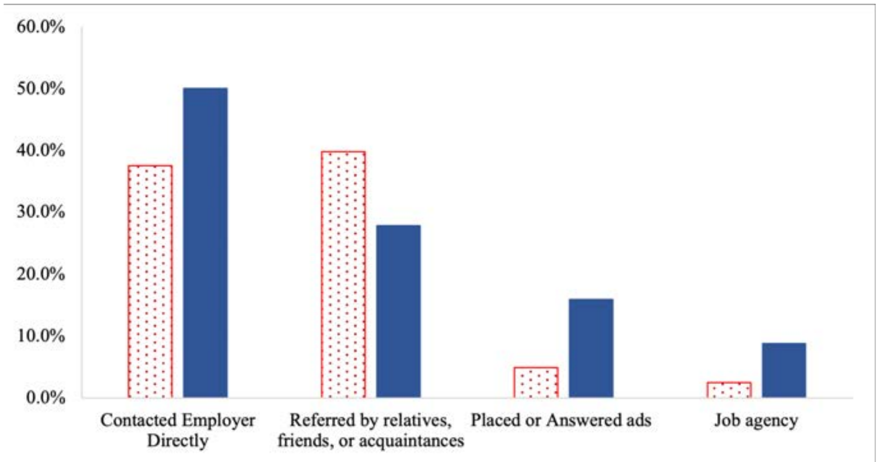
Zhang, J. and Wu, J. (2018). The chinese labor market, 2000–2016. *IZA World of Labor*, (437).

Zhao, Y. (2003). The role of migrant networks in labor migration: The case of china. *Contemporary Economic Policy*, 21(4):500–511.

Zhu, J. (2013). *Chinese SOEs and "invisible benefits"*. Retrieved May 13, 2013, from https://www.ft.com.

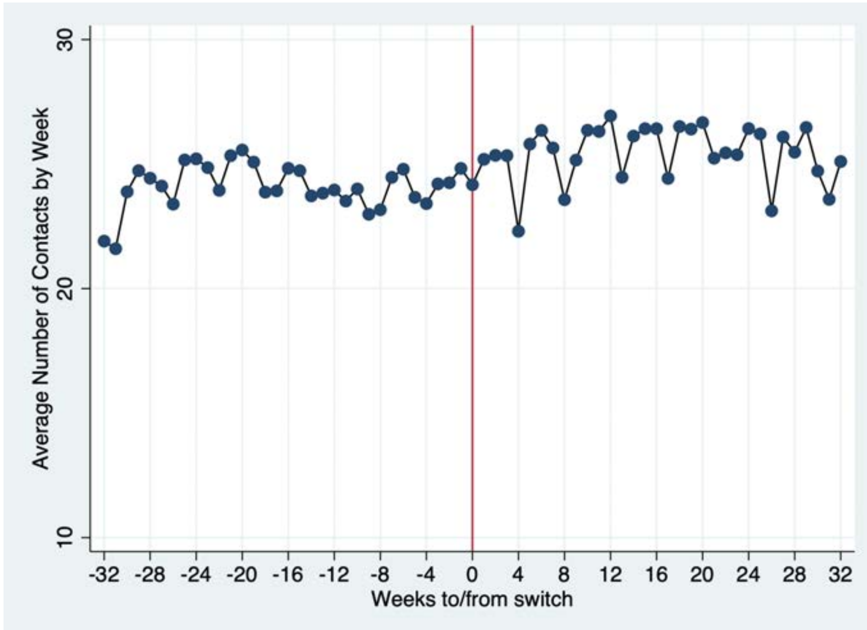
Figures and Tables

Figure 1: Job Search Methods in China vs. the United States (2014)



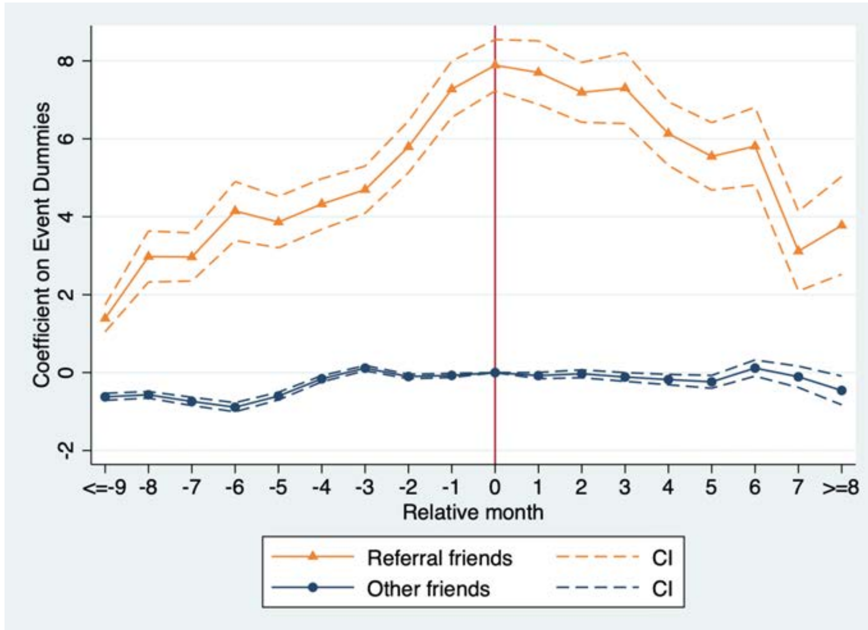
Notes: this figure shows the fraction of job seekers using different search methods to look for jobs. Red dotted (blue solid) bars represent China (United States). Source: the 2014 China Family Panel Studies and the 2014 U.S. Current Population Survey.

Figure 2: Job Switchers' Social Contacts Before and After Job Switch



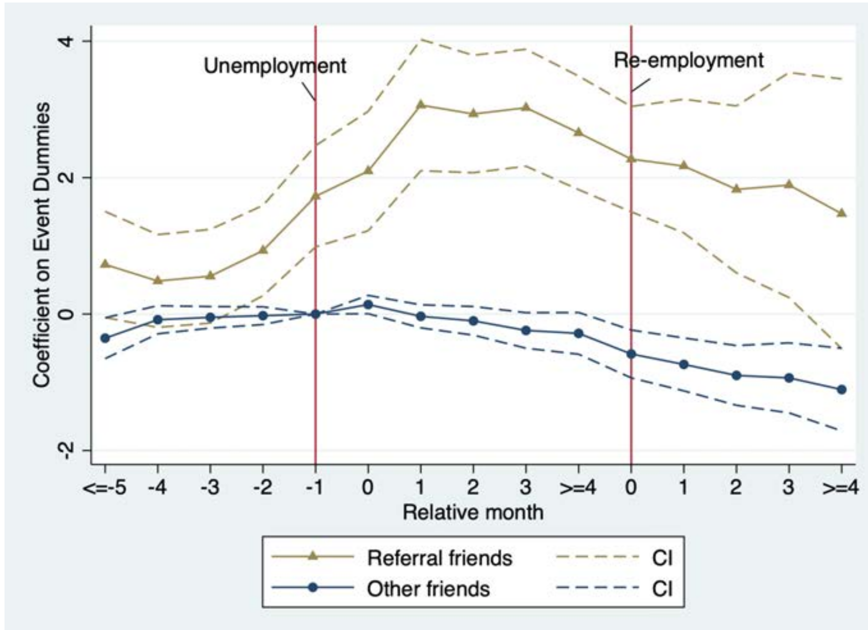
Notes: this figure plots the average number of social contacts who communicated with a job switcher in each week before and after the job switch. The vertical line indicates the week of job switch.

Figure 3: Calls to Referral vs. Nonreferral Friends



Notes: this figure plots the coefficient estimates and their 95% confidence intervals for an event study that examines the number of calls between a job switcher and his referral and nonreferral friends. The orange line (with triangles) represents calls between switchers and their referrals (nobs = 238,092). The blue line (with dots) represents calls between switchers and their nonreferral friends (nobs = 4,759,176). The vertical line indicates the month of job switch. The reference group is the call frequency between switchers and nonreferral friends one month prior to the job switch. The period prior to the ninth month before the job switch is binned with the ninth month, and the period after the eighth month post the job switch is binned with the eighth month. Switcher fixed effects and calendar month fixed effects are included in the regression.

Figure 4: Calls to Referral and Nonreferral Friends by Individuals with Unemployment Spells



Notes: this figure plots the coefficient estimates and their 95% confidence intervals for an event study that examines the number of calls between individuals with unemployment spells and their referral and nonreferral friends. The analysis limits to 3,638 individuals who became unemployed for at least eight weeks and then reemployed with valid work locations during our sample period. The brown line (with triangles) denotes calls between a reemployed individual and his referral friend (nobs = 53,244). The blue line (with dots) denotes calls between the re-employed individual and his nonreferral friends (nobs= 747,804). The first vertical line denotes the month immediately before the unemployment. The second vertical line denotes the month of reemployment. The period prior to the fifth month before unemployment is binned with month 5, the period after the fourth month post unemployment is binned with month 4, and the period after the fourth month post reemployment is binned with month 4. Calendar month fixed effects and individual fixed effects are controlled in the regression.

Table 1: Summary Statistics

(a) All users

	Mean	SD	N	2014 CFPS national survey Mean	SD
Female	0.36	0.48	435,098	0.45	0.50
Age 25-34	0.29	0.46	455,572	0.23	0.42
Age 35-44	0.26	0.44	455,572	0.24	0.43
Age 45-59	0.27	0.45	455,572	0.27	0.45
Age 60 and above	0.11	0.32	455,572	0.09	0.29
Age (midpoint)	40.18	11.97	435,194	39.28	14.07
Born in local province	0.75	0.43	455,572	0.76	0.43
Born in local city	0.39	0.49	455,572	-	-
Fraction of social contacts in Firm A	0.50	0.19	455,572	-	-
Job switcher	0.08	0.28	455,572	0.07	-

(b) Switchers vs. nonswitchers

	Nonswitchers			Switchers				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.36	0.48	398,742	0.36	0.48	36,356	-0.001	-0.45
Age (midpoint)	40.36	12.00	398,817	38.23	11.49	36,377	2.13***	32.49
Born in local province	0.75	0.43	417,470	0.74	0.44	38,102	0.01***	3.62
Born in local city	0.39	0.49	417,470	0.38	0.49	38,102	0.002	0.70
Fraction of social contacts in Firm A	0.50	0.19	417,470	0.51	0.19	38,102	-0.004	-0.53

Notes: the sample is restricted to 455,572 individuals with valid work information for at least 45 weeks during the sample period. Panel (a) reports summary statistics for all individuals. The last column presents the national average of the 2014 CFPS survey among individuals with phone-related expenses that exceed RMB 30 per month, weighted by the sampling weights. Panel (b) reports the summary statistics for switchers and nonswitchers separately, as well as the t-test statistics for the differences in sample means. "Fraction of social contacts in Firm A" is the fraction of an individual's social contacts who are Company A's subscribers. "Job switcher" is a dummy for individuals who changed jobs, based on the criteria described in the text. The CFPS sample's job switching rate is the "on-job switch rate" calculated by Nie and Sousa-Poza (2017). "Age (midpoint)" uses the midpoint of each age range. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Information Flow and Worker Flows

Dependent variable:	District pairs		Neighborhood pairs		Location pairs	
Worker flows (l, k)	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Regression analysis			
Information flow (l, k)	n.a.	0.003***	n.a.	0.0002***	n.a.	0.00005***
		(0.0004)		(1.95e-07)		(1.34e-05)
Observations	251	251	987,713	987,713	159,856,138	159,856,138
R-squared	0.745	0.971	0.024	0.164	0.002	0.042
Area l + Area k	Yes	Yes	Yes	Yes	Yes	Yes
			Panel B: Prediction exercise			
RMSE	66.610	26.251	0.366	0.178	0.012	0.011
MAPE	596.736%	115.267%	2.610%	1.011%	0.0099%	0.0098%

Notes: this table examines the relationship between information flow and worker flows. a) The unit of observation is a pair of administrative districts in Columns 1 and 2, a pair of neighborhoods in Columns 3 and 4, and a pair of locations in Columns 5 and 6. There are 23 administrative districts, 1406 neighborhoods, and 17,881 locations in the city. In Panel A, the dependent variable, "Worker flows (l, k)," is the total number of workers moving between area l and area k. "Information flow (l, k)" is the total number of calls between area l and k. Standard errors are reported in parentheses and two-way clustered by area l and area k. *** p<0.01, ** p<0.05, * p<0.1. b) Panel B conducts a prediction exercise whereby we run regressions using the first half of the sample and predict worker flows between area pairs during the second half of the sample, following the same specification as in the top panel. Then we compare the observed and predicted worker flows and report the root mean squared error (RMSE) and mean absolute percentage error (MAPE) of the prediction exercise. The prediction accuracy increases significantly in the even-numbered columns.

Table 3: Information Diversity and Worker Flows

Dependent variable: log inflow	(1)	(2)	(3)	(4)
Social entropy	0.96***			1.05***
	(0.33)			(0.38)
Spatial entropy		0.09		-0.35
		(0.28)		(0.31)
Income entropy			0.49**	0.37*
			(0.19)	(0.19)
Total call volume	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	6,161	6,161	6,161	6,161
R-squared	0.68	0.68	0.68	0.68
Neighborhood FE	Yes	Yes	Yes	Yes
Number of neighborhood FE	1,183	1,183	1,183	1,183

Notes: this table examines the importance of information diversity in explaining worker flows. The unit of observation is a location with at least five workers and five residents. "Log inflow" is the log number of people moving to a location. Social entropy, spatial entropy, and income entropy are normalized Shannon entropies as defined in Online Appendix Section S2. Total call volume is the total number of calls in thousands from or to a given location. Company A's subscribers in each location is controlled in all specifications. Standard errors are clustered by neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Effect of Friends on Job Location Choices

Dependent variable: Probability i switches to location l	(1)	Individuals with similar job opportunities nearby (2)	(3)	(4)	(5)
Friend	0.34***	0.35***	0.34***	0.34***	0.34***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Friend moved before the switch			0.07***		
			(0.03)		
Friend living but not working in i's new workplace				0.16***	
				(0.02)	
Friend of i's nonreferral friends					0.14***
					(0.01)
Observations	1,120,797	915,251	915,251	915,251	915,251
R-squared	0.14	0.12	0.12	0.13	0.13
Controls	Yes	Yes	Yes	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes	Yes	Yes
Number of neighborhood FE	20,811	16,468	16,468	16,468	16,468

Notes: this table examines the effect of referrals on job switchers' location choices. A unit of observation is a switcher-location pair. "Friend" is a dummy variable that equals one if switcher i has at least one friend working at a given location. Columns 2 to 5 limit to job switchers for whom there is at least one other location within the same neighborhood that has vacancy listings in the same occupation and salary range as the one taken. Column 3 compares referral friends (friends working in the new workplace) with friends who moved away prior to the job switch. Column 4 contrasts referral friends with friends living in the new workplace. Column 5 compares referral friends with second-degree social ties who also work in the new workplace (they are friends of switcher i's nonreferral friends but do not directly communicate with switcher i). All columns have the old-by-new work neighborhood-pair fixed effects. The controls include location amenities (the number of restaurants, roads, parking lots, and schools within a 500-meter radius), gender interacted with schools and parking lots, age group dummies interacted with number of restaurants, migrant dummy interacted with the number of roads, and individual i's number of social contacts interacted with all location attributes. Standard errors are clustered by the neighborhood pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Referral Effect and Information Asymmetry

Dependent variable: Probability i switches to location l	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.35***	0.33***	0.33***	0.33***	0.34***	0.32***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Friend×Distance(job1-job2)		0.002***				
		(0.0004)				
Friend×Distance(home-job2)			0.002***			
			(0.0003)			
Friend×Young				0.04***		
				(0.01)		
Friend×Rural to urban					0.32***	
					(0.05)	
Friend×Changing sector						0.21***
						(0.02)
Observations	915,251	915,251	915,251	915,251	915,251	915,251
R-squared	0.12	0.13	0.13	0.12	0.13	0.13
Old x new work neighborhood FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of neighborhood FE	16,468	16,468	16,468	16,468	16,468	16,468

Notes: this table uses the same specification as that in Column 2 of Table 4 and interacts the "Friend" dummy with measures that reflect the extent of information asymmetry. "Young" refers to switchers between 25 and 34 years old. "Rural to urban" flags switchers who move from the rural to urban part of the city. "Changing sector" is one if the switcher changes his sector. Columns 2–6 also control for the baseline level of the interacted variable. Standard errors are clustered by the neighborhood pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Referral Benefits to Workers

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Wage at new job	ΔCoworker HP	PT to FT	Shorter commute	Non-SOE to SOE
Friend	0.62**	0.07*	0.014**	0.09***	0.012**
	(0.31)	(0.04)	(0.007)	(0.01)	(0.005)
Observations	17,615	23,323	19,431	29,117	15,881
R-squared	0.79	0.53	0.11	0.12	0.56
Residence Neighborhood FE	Yes	Yes	Yes	Yes	Yes
New work neighborhood FE	Yes	Yes	Yes	Yes	Yes

Notes: this table examines whether jobs through referrals provide higher (financial and nonfinancial) benefits to workers. The sample size varies due to missing observations. All regressions include the same demographic controls as in Table 4. "Wage at new job" is the annual payroll per worker in thousand RMB, weighted by employee sizes among firms in the new work location. "ΔCoworker HP" is the difference between coworkers' average house price (thousand RMB per m^2) in the new workplace and that in the old workplace. "PT to FT" is a dummy that equals one if the switcher works part time (30 hours or less per week) before the job change and full time (more than 30 hours) afterward. "Shorter commute" equals one if the commuting distance at the new workplace is shorter than before. "Non-SOE to SOE" is a dummy that equals one if the new workplace is an SOE dominant location (with the majority of employees working in SOE firms), while the previous location is not. Standard errors are two-way clustered by the residential neighborhood and new work neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Referral Benefits to Firms

Dependent variable: Panel A: Log of net inflow	(1)	(2)	(3)	(4)
Referral	0.60***	0.62***	0.62***	0.63***
	(0.13)	(0.14)	(0.14)	(0.14)
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.64	0.65	0.65	0.66
Panel B: Log of matching rate	(5)	(6)	(7)	(8)
Referral	0.91***	0.88***	0.88***	0.86***
	(0.24)	(0.26)	(0.26)	(0.27)
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.85	0.87	0.87	0.87
Panel C: Log of firm growth rate	(9)	(10)	(11)	(12)
Referral	0.49***	0.45***	0.44***	0.45***
	(0.11)	(0.10)	(0.10)	(0.11)
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.76	0.83	0.83	0.83
Controls				
Firm attributes	No	Yes	Yes	Yes
Previous growth rate	No	No	Yes	Yes
Employee attributes	No	No	No	Yes
Neighborhood FE	Yes	Yes	Yes	Yes

Notes: this table examines whether firms benefit from referred jobs. A unit of observation is a location that has at least one firm with more than 100 employees and positive hiring. We report a range of the observation number to keep Company A anonymous. The number of neighborhood fixed effects is 225. “Referral” takes value one if at least one switcher joining the firm has a friend working in the firm. “Net inflow” is the number of hires minus separations. “Matching rate” is defined as the net inflow over the number of vacancies. “Firm growth rate” is measured as the net inflow over the employee size. Firm attributes include age, industry, SOE dummy, average employee size, real capital, and employee growth rate from 2010 to 2015. Employee attributes includes share of female and migrants and the average age of preexisting employees. Firm network size, measured by the number of distinct contacts of the firm’s preexisting employees, as well as the number of Company A’s users at each location is controlled in all columns. Standard errors are clustered by neighborhood and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8: Referral Effect and Robustness to Local Labor Demand

Dependent variable: Probability i switches to location l	(1)	(2)	(3)	(4)	(5)
Friend	0.34***	0.34***	0.34***	0.34***	0.33***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Friend moved before the switch	0.07***	0.07***	0.07***	0.07***	0.06***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Location-level controls and interactions with demographics					
Number of vacancies in each occupation	Yes	Yes	Yes	Yes	Yes
Firm characteristics	No	Yes	Yes	Yes	Yes
Number of calls by working population	No	No	Yes	Yes	Yes
Number of calls by residential population	No	No	No	Yes	Yes
Turnovers before switch/Average turnovers	No	No	No	No	Yes
Observations	915,251	915,251	915,251	915,251	915,251
R-squared	0.125	0.125	0.127	0.127	0.137
Old x new work neighborhood FE	Yes	Yes	Yes	Yes	Yes
Number of neighborhood FE	16,468	16,468	16,468	16,468	16,468

Notes: this table replicates Column 3 of Table 4 and examines robustness to local labor market demand. Column 1 includes the number of vacancies by occupation within a 500-meter radius of each location and the interactions with switcher i's demographics (gender, age group and migrant dummies). Column 2 adds firm characteristics (age, industry, number of employees, average wage, revenue, and real capital from 2011 to 2015) and the interactions with i's demographics. Column 3 also controls for the total number of calls associated with the working population (excluding switcher i) in each location. Column 4 further controls for the call volume associated with the residential population. Column 5 controls for potential changes in location l's work environment prior to the job switch by adding as a regressor the ratio of the average monthly number of people leaving location l during the three months prior to the job change over the average monthly turnovers at location l for the rest of the sample period. We add a dummy for locations with no switchers. Standard errors are clustered by the neighborhood pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Referral Effect and Robustness to Homophily

Dependent variable: Probability i switches to location l	(1)	(2)	(3)	(4) $K=5$	(5) $K=10$	(6) $K=100$
Friend	0.32***	0.34***	0.33***	0.34***	0.34***	0.34***
	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)
Homophily controls						
Whether friend at l has same-gender, age group, birth county and house price as i	Yes	Yes	Yes	No	No	No
Share of mutual friends and neighborhoods between i and friend at l	No	Yes	Yes	No	No	No
Difference between the demographics of i's friends and demographics of friend at l's friends	No	No	Yes	No	No	No
Dummy for 'i and friend at l in the same cluster' (K is the number of clusters)	No	No	No	Yes	Yes	Yes
Observations	915,251	915,251	915,251	915,251	915,251	915,251
R-squared	0.12	0.12	0.12	0.12	0.12	0.12
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of neighborhood FE	16,468	16,468	16,468	16,468	16,468	16,468

Notes: this table examines referral estimates' robustness to homophily. The columns use the same specification as that in Column 2 of Table 4 but add more controls for homophily. Column 1 controls for whether the friend at location l (referral friend) has the same gender, age group, birth county, and similar house price as switcher i. Column 2 further adds the share of mutual friends as defined by the Jaccard index between i and the referral friend. It also examines the geographical coverage of i's friends and referral friend's friends by including the share of overlapping work neighborhoods (also defined by the Jaccard index) between i's friends and the referral friend's friends. Column 3 first calculates the average demographics of i's social contacts: the fraction of i's friends who are female, migrants, in each birth county and age group, and their average house price. Then it obtains the average demographics of referral friend's social contacts and uses the demographic differences between i's social contacts and referral friend's social contacts as controls. In locations with more than one friend, friend attributes are constructed using the average. In Columns 4 to 6, we group switchers and their friends into clusters using the k-means clustering algorithm based on each individual's own characteristics and their social ties' characteristics. Standard errors are clustered by the neighborhood pair in all columns and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Referral Effect and Robustness to Preference for Working with Friends

Dependent variable: Probability i switches to location l	(1)	(2)	(3)	(4)
Friend	0.34***	0.35***	0.31***	0.35***
	(0.01)	(0.01)	(0.02)	(0.01)
Number of friends at l	0.005		0.005	
	(0.004)		(0.004)	
Dummy for Number of friends at $l \geq$ Number of friends at old place		0.001***		0.002**
		(0.0004)		(0.001)
Observations	915,251	915,251	915,251	915,251
R-squared	0.12	0.12	0.13	0.12
Controls	Yes	Yes	Yes	Yes
Friend socio-demographic controls	No	No	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes	Yes
Number of neighborhood FE	16,468	16,468	16,468	16,468

Notes: this table examines the referral estimate's robustness to the possible presence of preferences to work with friends. The columns use the same specification as that in Column 2 of Table 4. Column 1 additionally controls for the number of friends at location l. Column 2 includes a dummy that takes value one if the number of friends at location l is equal to or larger than that at the old workplace. Column 3 replicates Column 1 but adds controls for the fraction of friends with the same gender, age group, birth county, and similar house price as switcher i. Column 4 replicates Column 2 and includes dummies on whether the number of friends with the same gender, age group, birth county, and similar house price at location l is equal to or larger than that at the old workplace. Standard errors are clustered by the neighborhood pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Comparison with Referral Proxies in the Literature

Dependent variable: Probability i switches to location l	(1)	(2)	(3)	(4)
Friend definition				
Residential neighbor	0.21***		0.18***	
	(0.01)		(0.01)	
Same birth county		0.10***		0.09***
		(0.01)		(0.01)
Friend, not neighbor			0.25***	
			(0.01)	
Friend, not same birth county				0.36***
				(0.02)
Observations	915,251	915,251	915,251	915,251
R-squared	0.15	0.10	0.19	0.15
Controls	Yes	Yes	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes	Yes
Number of work neighborhood pair FE	16,468	16,468	16,468	16,468
Residential neighborhood FE	Yes	No	Yes	No
Number of residential neighborhood FE	1,067	NA	1,067	NA
Birth county FE	No	Yes	No	Yes
Number of birth county FE	NA	17	NA	17

Notes: this table uses the same specification as that in Column 2 of Table 4 and contrasts our referral measure with proxies in the literature. "Residential neighbor" is a dummy that equals one if switcher i has at least one residential neighbor who works in the new work location. "Same birth county" is a dummy that equals one if there is at least one individual who works in switcher i's new work location and shares the same birth county as switcher i. Old-by-new work neighborhood-pair fixed effects are included in all columns. Columns 1 and 3 also control for residential neighborhood fixed effects. Columns 2 and 4 control for the birth county fixed effects. Standard errors are reported in parentheses and clustered by the old-by-new work neighborhood-pair as well as the residential neighborhood in Columns 1 and 3 and by the old-by-new work neighborhood-pair as well as the birth county in Columns 2 and 4. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Phone Calls, Internet Browsing, and Data Usage

	All users			Switchers		
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Average monthly calls						
Whether 4G	66.85***	28.36***	3.85***	58.59***	28.13***	5.78**
	(1.22)	(0.83)	(0.78)	(3.26)	(2.41)	(2.34)
Observations	350,496	341,840	341,840	30,299	29,539	29,539
Panel B: Average monthly calls						
Data volume per month (GB)	63.75***	40.60***	26.12***	14.65***	7.03***	1.45**
	(1.28)	(1.18)	(1.45)	(0.82)	(0.60)	(0.59)
Observations	384,644	375,120	375,120	32,641	31,821	31,821
Panel C: Number of calls in the same week as the browsing data						
Browsing time (in thou. minutes)	15.67***	3.91***	2.02***	16.09***	4.22***	3.48***
	(0.98)	(0.53)	(0.44)	(1.31)	(1.12)	(1.10)
Observations	316,976	308,977	281,225	28,348	27,641	25,297
Residential neighborhood FE	Yes	Yes	Yes	Yes	Yes	Yes
Social demographic	No	Yes	Yes	No	Yes	Yes
Other controls	No	No	Yes	No	No	Yes

Notes: this table examines whether consumers who make more phone calls also use more intensively other information channels (Internet and mobile apps as proxied by the 4G network). Each cell denotes a separate regression. Columns 1–3 include all users with 45 weeks of valid work information and nonmissing residential locations. Columns 4–6 limit to job switchers. Panel (a) regresses an individual's average monthly calls on whether his phone device is compatible with the 4G network. Panel (b) regresses an individual's average monthly calls on his cellphone plan's Internet data volume. Panel (c) regresses an individual's number of calls during the second week of May 2017 on his Internet browsing time (measured in thousand minutes) in the same week. All columns control for residential neighborhood fixed effects. Columns 2 and 5 control for demographics, including age, gender, whether born in the city, and the number of contacts. Columns 3 and 6 additionally control for phone price, monthly fees, number of weeks in our sample, and average working hours per week. Standard errors are clustered at the residential neighborhood level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.