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

We thank Susan Athey, Patrick Bayer, Gilles Duranton, Giacomo De Giorgi, Jessie Handbury, Tatsiramos Konstantinos, Mike Lovenheim, Michele Pellizzari, and Steve Ross for helpful comments. 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
JEL No. J60,R23

ABSTRACT

Information is a crucial ingredient in economic decision making. Yet measuring the extent of information exchange among individuals and its effect on economic outcomes is a difficult task. We use the universe of de-identified cellphone usage records from more than one million users in a Chinese city over twelve months to quantify information exchange among individuals and examine the role of referrals – human carriers of information – in urban labor markets. We present the first evidence that information flow (measured by call volume) correlates strongly with worker flows, a pattern that persists at different levels of geographic aggregation. Condition on information flow, socioeconomic diversity in information sources (social contacts), especially that associated with the working population, is crucial and helps to predict worker flows. We supplement our phone records with auxiliary data sets on residential housing prices, job postings, and firm attributes from administrative data. Information passed on through referrals is valuable: 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 sector. Firms receiving referrals are more likely to have successful recruits and experience 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 (International Food Policy
Research Institute)
Senior Fellow
Washington, DC 20006-1002
Y.Liu@cgiar.org

Eleonora Patacchini
Department of Economics
Cornell University
430 Uris Hall
Ithaca, NY 14853
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. In classical analysis, it is assumed that agents choose actions to maximize payoff under *perfect* information (Arrow and Debreu, 1954). In reality, information is rarely perfect. Agents' information set differs substantially, as highlighted by the influential literature on information asymmetry (Vickrey, 1961; Mirrlees, 1971; Akerlof, 1970; Spence, 1973; Rothschild and Stiglitz, 1976). In addition, information exchange and acquisition is costly and crucially depends on the interaction between information carrier and the receiving parties (agents, institutions, etc.).

Quantifying the effect of information exchange among social entities and individuals on economic outcomes is a difficult task, because it is challenging to empirically 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 technologies and Global Positioning System (GPS) in mobile phone devices provides a novel avenue that allows researchers to quantify the extent of information exchange among individuals, while also tracking their movements in the physical space. Datasets derived from geocoded phone communication records present three unique advantages over existing data sets. First, the frequency and intensity of calling 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 portrays a more accurate profile of individuals' social network than surveys that are commonly used in the literature. Existing research have documented that mobile phone usage accurately predicts human mobility (Gonzalez et al., 2008), poverty and wealth (Blumenstock et al., 2015; Blumenstock, 2018), credit repayment (Bjorkegren and Grissen, 2018), restaurant choice (Athey et al., 2018), and residential location choice (Buchel et al., 2019).

In this paper, we focus on the effect of information on the dynamics of the labor market, arguably one of the most important markets in a society. Our empirical research has the following goals. First, we investigate to what extent information flows are accompanied by worker flows. Second, we examine how information flow via human carriers – friend referrals – affects job transitions and worker-vacancy match efficiency.

Toward this end, we exploit the universe of de-identified cellphone usage records from all users in a large Chinese city served by a major telecommunication service provider over twelve months. These detailed records enable us to construct measures of information exchange between geographic areas and among individuals and crucially, variables on each user's employment status, history of work locations, home locations, as well demographic attributes (gender, age, and birthplace). We supplement our phone records with auxiliary data sets on residential housing prices, job postings, and firm attributes (industry and payroll) for additional socioeconomic measures.

We proceed in several steps. First, we provide the first empirical evidence 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. Conditioning on the amount of information exchanged, diversity of individuals' social contacts (sources of information) also matters.[1] 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 information sources possessed by the working population is far more critical than that by the residential population. Surprisingly, our data exhibit remarkable similarity to the UK data analyzed in Eagle et al. (2010) in terms of the relationship between information diversity and economic development, highlighting the potentially wide applicability of this finding in different settings.

Having documented the importance of information flow with respect to worker flow, we examine the role of referrals – human carrier of information – on job switch. When individuals move to a pre-existing friend's workplace, such a friend is defined as a referral. We first document that the intensity of information flow between workers and their referrals exhibits an inverted-U shape that peaks at the time of job switch. In contrast, the information flow between workers and non-referral friends remained stable throughout the sample period, with no noticeable differences in the months leading to the job switch. The distinctions in calling patterns are not driven by changing numbers of social contacts, which remain steady throughout our sample period. These patterns provide suggestive evidence that individuals seek valuable job-related information from their social contacts.

One might be concerned that the referral definition in our sample suffers from several confounding factors. First, firms sometimes relocate, consolidate, or open new plants in different areas. If employees are relocated in different time periods, we might observe flows of workers moving to the work location of pre-existing social contacts. We tackle this problem by adding the *interaction* of the origin and destination neighborhood fixed effects. Essentially, we compare individuals with the same origin-destination neighborhood pair but different social networks and examine their choice of workplace locations with or without friends.

The second confounding factor, a long-standing challenge in the literature using observational data, is the difficulty to distinguish 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. We leverage on the richness and structure of our data and conduct a battery of tests. First, we limit our analysis to individuals for whom there is at least another location within the same

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

neighborhood that has vacancy listings in the same occupation as the one that the mover takes. This mitigates the concern that individuals sort into friends' locations that provide the only appropriate employment opportunity in the area.

Second, we distinguish between friends who are currently working in the location and friends who used to work there but have moved away prior to the job switch. Given that sorting into friendship 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 definition of referrals primarily reflects sorting. Third, we compare friends who work vs. live at a location. Sorting would imply a similar effect for these two types of friends. On the other hand, larger estimates for friends working in the location would be consistent with referrals: affiliation with the workplace enable friends working there to have an information advantage of job openings. Our results illustrate that friends currently working in the location is much more important that friends who have moved away prior to the job switch or who live there. In all these tests, we find evidence supportive of our definition of referrals who pass on valuable work-related information to their social contacts.

According to our analysis, one out of four jobs are based on referrals. Having a referral in a location increases the likelihood that an individual moves there by four times, a pattern that is robust across a host of specifications and consistent with previous studies in various countries (Ioannides and Loury, 2004). Referrals are more effective for young workers, people switching jobs from suburbs to the inner city, and people who change sectors. These results are in line with the observation that information asymmetries are more severe in these settings.

Job information passed on via referrals is valuable for workers. Specifically, referral jobs are associated with higher wage and non-wage benefits, shorter commutes, and a higher likelihood to transition from part-time to full-time and from regular jobs to premium ones. Information transmitted through the referral networks is also valuable for firms. Firms whose employees have a larger social network are more likely to have successful recruits, achieve higher retention rates, and experience faster growth. In addition, we find suggestive evidence that referrals improve labor market efficiency by providing better matches between workers and vacancies and that referrals improve labor market inequality, as females and migrants are more likely to find jobs through referrals.

Our work contributes to the emerging literature demonstrating how the widespread use of electronic technologies, and consequently the wealth of information on individual (or firm) digital fingerprints, opens new frontiers for urban economics (Glaeser et al., 2015; Donaldson and Storeygard, 2016). A pioneering study by Henderson et al. (2012) exploits satellite data to improve the analysis on urban economic activities at 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 that utilize such data to improve our understanding of urban development. Other studies include housing decisions (Bailey et al., 2018), households' responses to income shocks (Baker, 2018), entrepreneurship and capital investment (Jeffers, 2018). Our work contributes to this literature by combining mobile phone records with traditional socioeconomic data in creative ways to shed light on urban labor market mobility at a fine geographic and temporal level.

Our work is related to the empirical literature on information economics. Recent studies have shown increasing information transparency (for example, through better labels and postings) helps consumers' perception of product attributes (e.g., Smith and Johnson 1988), improves consumer choices (e.g., Hastings and Weinstein 2008), 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. Our study is also related to the literature on diversity, including Page (2007) and Eagle et al. (2010). We propose novel measures on the diversity of socioeconomic outcomes and illustrate their importance in shaping worker flows.

Another relevant strand of literature examines the role of social network in job search. This literature identifies referred workers using surveys or assuming interactions and exchange of job information between social ties, such as fellow workers, family ties, ethnic groups, residential neighbors, and facebook friends.[2] The paper closest to ours is Bayer et al. (2008) that also studies the importance of referral effects in an urban market. Using Census data on residential and employment locations, they document that individuals residing in the same city block are more likely to work together than those in nearby blocks, and interpret these findings as evidence of social interactions. We contribute to this literature by providing a superior measure of social network and information exchange among individuals and bringing complementary data on vacancies and firm attributes that cover a diverse set of economic outcomes.

Our investigation is important for three reasons. First, despite the importance of social network in job transitions (Topa, 2011; Schumutte, 2016), accurately measuring these networks has remained a challenge. Our analysis illustrates how the availability of longitudinal and geocoded call records is an important step towards overcoming the measurement challenge. Second, these new types of data typically lack information on socioeconomic attributes. Our analysis highlights the potentials of supplementing these data with traditional sources that could significantly broaden the sets of questions to be analyzed. Third, our analysis indicates that information is an important pathway for understanding labor market dynamics. In addition, diversity in information sources can be crucial in improving economic outcomes.

[2]The existing literature has proposed several proxies for social network, such as former fellow workers (Cingano and Rosolia 2012; Giltz 2017; Saygin et al. 2018), family ties (Kramarz and Skans 2014), individuals belonging 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; Schumutte 2015), and Facebook friends (Gee et al. 2017)

The paper proceed as follows. Section 2 presents motivating evidence that information flow strongly correlates with flow of workers. Section 3 discusses data and institutional backgrounds. Section 4 illustrates the model and reports results from the core empirical analysis. Section 5 concludes.

2 Motivating Evidence: Information Flow and Worker Flow

We are interested in understanding how information exchange affects urban labor markets. Given the challenges in empirically measuring the extent of information flow across geographic regions and social groups, we resort to non-standard datasets. Our analysis is made possible by a unique data set that contains the universe of phone records for all subscribers in a metropolitan city by a major telecommunication operator in China. This data set provides a superior coverage on individuals' social network and allows us to identify their geocoded work place and residence (see Section 3 for details.) We use the number of phone calls between two areas to measure information flow and relate it with worker flow that is constructed using the same data.[3] To our best knowledge, this is the first empirical analysis that examines the empirical relationship between information flow and worker flow.

Descriptive Evidence We document a strong correlation between information flow as measured by the number of phone calls and worker flow between pairs of geographical locations at varying level of spatial aggregation. At the highest level, the city is divided into twenty-three administrative districts. These administrative districts are further broken into 1,406 neighborhoods that are delineated by major road (or 'cells'), with each neighborhood populated by a varying number of locations.[4] There are close to eighteen thousand locations in total.

To illustrate the patterns of worker flow and mobile communication, we first plot in Figure 1 worker flows against the number of calls between a pair of administrative districts for ten randomly chosen districts within the city proper. Blue non-directional edges correspond to the number of job switches among the relevant pairs, with the width of each edge scaled proportionately to the number of switches. Red non-directional edges denote the average number of weekly calls, with a scaled edge-width as well.[5] Note the remarkably strong correlation between

[3] An alternative measure of information flow is the total call volume in minutes. This measure delivers similar results.

[4] The average size for an administrative district, a neighborhood (cell) in the city proper, and a neighborhood (cell) in the suburb is 712 km^2, 0.45 km^2, and 25.03 km^2, respectively. A location is a building complex within a neighborhood (cell).

[5] The graph is produced using the Fruchterman & Reingold algorithm that aims to distribute vertices evenly (Fruchterman and Reingold 1991).

the two types of edges. City districts with frequent information exchange (blue lines) also have more worker flows (red lines), with the correlation between these two series exceeding 0.94. The two nodes with the thickest edges are a commercial center of the city with large retail chains and an urban core with the second highest GDP among all areas in the city. The strong correlation remains when we zoom out to the entire city, including districts in suburbs with less economic activity (fewer job switchers and lower call volumes).

Some correlation naturally arises from uncontrolled geographical attributes, like the example above of two economic centers. To address this, we run a regression analysis and control for origin and destination fixed effects. Regressing the worker flow between an district-pair on the total number of phone calls for the same pair leads to a significant (both statistically and economically) coefficient: three hundred more calls are associated with one more job switch (Column 1 in Table 1a). Using a log-log specification suggests that doubling the number of calls is associated with a 35% increase in worker flows.[6] The R-squared is 0.24 when the number of calls is the only regressor and jumps to 0.90 when origin and destination fixed effects are included.

A key premise of our analysis is that call volumes serve as a proxy for the amount of information exchanged between individuals. To better measure job-related information that facilitates worker flows, we limit to calls received or made by job switchers prior to the job change (Column 2). In practice, some calls might be initiated after having decided to move and reflect communications arising from newly established (work) relationships. In Column 3 and 4, we thus further exclude calls made within one month (Column 3) and three months (Column 4) of the job switch. When excluding calls that may be unrelated to job-openings, the magnitude strengthens as we move from Column 2 to Column 4, with one additional worker flow following eight more calls (Column 4).

When we zoom into finer geographical levels, this strong correlation persists. Table 1b presents coefficient estimates when we regress worker flow on information flow at the location-pair level. Our data cover eighteen thousand locations and millions of location pairs. Predicting the exact location (a building complex in our example) of job movers is a very demanding exercise. Reassuringly, the positive correlation between information and worker flow maintains even at this fine level, with one thousand more calls associated with one additional worker flow using the switcher sample (Column 4). At the neighborhood level (an geographical area in between the administrative district and a location), the correlation between information flow and worker flow is 0.75. Regressions using neighborhood observations deliver very similar results, corroborating the importance of job-related information flow in facilitating worker flows.

[6] Results are available upon request.

Out-of-sample Prediction Existing studies have shown that mobile phone usage can predict economic activities. Here we follow Kreindler and Miyauchi (2019) and use the relationship between information exchange and worker flow that is uncovered using the first half of the sample from November 2016 to April 2017 to predict worker flow in the second half of the sample from May 2017 to October 2017.

Specifically, we split the sample of job switchers into two groups, Group A and B, that include workers who changed jobs in the first and second half of the sample, respectively. The information we have includes mobile communication between every neighborhoods pair i, j. Our goal is to use Group A as a training sample to predict worker flow between a neighborhood pair. Once we obtain our prediction for Group B, we examine the accuracy of the prediction. We first check the correlation between our prediction and the observed outcome. The closer this correlation is to one, the better. Then we regress the observed outcome on predicted value and report the R-squared values. We use several models: a) a linear model with neighborhood fixed effects; b) a linear spline model; and c) a cubic-spline model.[7]

As shown in Table 2 (where even columns also control for cell fixed effects), the out-of-sample prediction exercise does well. In all cases we examined, the correlation between our prediction and the observed outcome for the target sample is close to one, varying between 0.97 to 1.03 depending on specifications. The R-squared is 0.31, which is high for cross-sectional studies with a large sample. These results are encouraging and suggest that information flow is an important predictor of worker flow.

Diversity and Economic Outcome The results above provide evidence of a strong parallel movement between information flow and worker flow. Both the sociology and economic 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. Economic opportunities are diverse and more likely to come from contacts outside a tightly knit local friendship group. 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 different social entities.

Following Eagle et al. (2010), we define three diversity measures using the normalized Shannon entropy: social entropy, spacial entropy, and income entropy.[8] The social entropy measures

[7]We use the default number of spline knots in STATA.

[8]Cover and Thomas (2006) is a classic textbook on information theory and entropy measures.

the diversity of individuals' social ties and is defined as:

$$\begin{aligned} D^{\text{social}}(i) &= -\frac{\sum_j P_{ij} log(P_{ij})}{log(\text{NumFriend}_i)} \\ &= -\frac{\sum_j \frac{\nu_{ij}}{V_i} log(\frac{\nu_{ij}}{V_i})}{log(\text{NumFriend}_i)} \end{aligned}$$

where P_{ij} is the probability of communication between individuals i and j and is measured by the ratio of ν_{ij}, the number of calls between i and j, and V_i, the total number of calls placed or received by i. The denominator, log of total number of friends that individual i has, is a scaling number that normalizes the Shannon entropy. Normalized Shannon entropy measures are guaranteed to vary from zero and one and are comparable across different measures, with higher values representing more diverse outcomes.

Spatial entropy measures the diversity of individuals' social ties in geographic locations:

$$\begin{aligned} D^{\text{spatial}}(i) &= -\frac{\sum_l P_{il} log(P_{il})}{log(\text{NumLocation}_i)} \\ &= -\frac{\sum_l \frac{\nu_{il}}{V_i} log(\frac{\nu_{il}}{V_i})}{log(\text{NumLocation}_i)} \end{aligned}$$

where P_{il} is the probability of communication between individuals i and location l, ν_{il} is number of calls between i and location l, and V_i is defined as above. The denominator $log(\text{NumLocation}_i)$ is the log number of locations where i has social contacts.

Finally, we define the income entropy as:

$$\begin{aligned} D^{\text{income}}(i) &= -\frac{\sum_d P_{id} log(P_{id})}{log(\text{NumDecile}_i)} \\ &= -\frac{\sum_d \frac{\nu_{id}}{V_i} log(\frac{\nu_{id}}{V_i})}{log(\text{NumDecile}_i)} \end{aligned}$$

where ν_{id} is the number of calls between i and all individuals whose housing price falls in the dth decile of the overall housing price distribution. The variable V_i is as defined above. Similar to the other two entropy measures, the normalization is through the number of unique deciles that are spanned by the housing prices of individual i's friends. Income entropy measures socioeconomic diversity among i's social network.

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 residing or working in each location. A high value indicates that the working or residential population at a particular location communicates with a diverse set of information sources. To examine the importance of diversity, we regress the log of worker flow on the average entropy measures at the location level. The control variables include total call volumes, which as shown above is an important predictor of worker flows. We also include the number of individuals (subscribers of Company A) observed in a location to control for the 'scale' effect: more populated areas naturally have a higher job inflow. Finally, all regressions include neighborhood fixed effects. Hence our key parameters are estimated from within-neighborhood across-location variation.

Columns 1 to 3 of Table 3 include one entropy measure at a time, while Column 4 includes all three measures.[9] Both social and income entropy measures, which reflect the socioeconomic diversity of individuals' information sources, have a sizable and significant impact on the job inflow, *conditional on* the total number of calls. One standard deviation increase in social and income entropy raises the worker inflow by 3% and 10%, respectively, which is remarkable. Spatial entropy, on the other hand, is insignificant with a negative sign. This might reflect that our sample constitutes individuals from the same city with limited spatial diversity. Among these three measures, social and income entropy appear to be equally important in affecting worker flows, though the correlation between income entropy and worker flows is stronger.

Next we examine the relative importance of information acquired by the *working* vs. *residential* population. Table 4 repeats the regressions in Table 3 but includes the entropy measures for both the working and residential population. As shown in Appendix Table A1, the entropy measures for these two populations have similar distributions. However, information diversity of the working population has a much stronger correlation with worker flows than that of residents in the same location. Once we condition on the entropy measure of the working population, the coefficient associated with the residential population is insignificant and much smaller. While our analysis is descriptive, these results highlight the heterogeneous values in information possessed by different social groups and reflect the fact that information about jobs is predominantly in the domain of the working population.

It is worth noting that our results reveal remarkable similarity to the findings in Eagle et al. (2010) that examines phone calls in UK in 2005 and relates communication flows to the socioeconomic well-being of communities. While the average number of monthly contacts is higher in our context, 24 vs. 10.1 in UK, which reflects a denser social network in China, the average minimum number of direct or indirect edges connecting two individuals is very similar (10.4 in our context vs. 9.4 in UK). In addition, as in our setting, there is a surprisingly strong correlation (varying from 0.58 to 0.73) between information diversity and socialeconomic development of different communities in UK. These results reflect common features of the role of information (diversity) that are at play across a diverse set of socioeconomic contexts and

[9]The coverage for suburb locations is sparse. Here we limit to locations with at least five workers and five residents. Results are similar if we use all locations or limit to those with at least ten observed workers and residents.

are not unique to specific markets or time periods.

Having illustrated the high correlation between information exchange and job flows, we next turn to the bulk of our empirical analysis where we focus on a specific channel of information at work: referrals, the human carrier of information on job openings. The existing literature has documented that referrals are extremely important in facilitating the match between workers and jobs. In fact, 30 to 60 percent of all jobs are typically found through informal contacts rather than formal search methods (Topa, 2001; Burks et al., 2015), a universal pattern that holds across countries, over time, and regardless of occupation or industry. Despite the importance of referrals, our understanding regarding how the information exchange between referrals and job seekers affects the performance of labor markets is limited because of the scant information on who actually interacts with whom. Our unique calling data provide a superior coverage on information flow through social networks. In addition, we bring in a large number of complementary data sets with rich measures on vacancies, firm attributes, and job amenities. In the following, Section 3 describes in detail various data sets assembled in this paper and Section 4 presents the core empirical analysis.

3 Data and Institutional Background

We have compiled a large number of data sets for our analysis. Besides data on phone records, we have put together axillary data sets on residential housing prices, vacancies (job posting data), and firm attributes (administrative datasets on firms registered in the same city).

Call Data Our anonymized call data consist of the universe phone records for 1.6 million mobile-phone users in a northern city in China from November 2016 to October 2017. The data provider is a major telecommunication operator and mobile service provider in China (hereafter Company A), which serves between 30-65% of all mobile phone users in the relevant city.[10] Our data set contains information on all of company A's subscribers in the city we study.

Cellphone usage records are automatically collected when individuals send a text message, make a call, or browse the internet. These records include identifiers (IDs), the user location at the time of usage, time and duration of usage. There are several advantages of this data set. First, unlike social media, cellphone penetration rate is very high in China. According to the China Family Panel Studies (hereafter CFPS), a nationally representative annual longitudinal survey on individuals' social and economic status since 2010, 85% of correspondents sixteen years and older report possessing a cellphone. Second, our data include the de-identified ids of

[10]There are three major mobile service providers in China. We report a wide range of the market share to help keeping company A anonymous.

the calling party and the receiving party, weekly call frequency and call duration in seconds, and whether or not a user is company A's subscriber.

Third, the data set provides demographic information of the subscribers, such as age, gender, and place of birth. The birth county enables us to classify users into migrants and non-migrants, which is analogous to ethnicity groups commonly used in the literature. Social connections (or *guan-xi*), whether from the same region, being friends or relatives, play an important role in China's labor market. For example, migrants are much more likely to refer and work with other migrants from their birth city and province (Dai et al., 2018).

Fourth, and most importantly, our data set contains information on users' phone usage and geocoded locations at the time of use. When users activate their mobile devices by making phone calls, sending messages, or browsing the Internet, the serving tower station records a geographical position in terms of the longitude and latitude for the device, which is accurate up to a 100-200 meter radius, roughly the size of a large building complex. Our data vendor provided us each individual's primary work location in a given week, defined as the location with the most frequent phone usage between 9am and 6pm during the weekdays, as well as a residential location (the location with the most frequent phone usage between 10pm and 7am) for the same week.[11] These recorded locations trace out individuals' (as well as their friends') spatial trajectory over time and allow us to construct diverse types of social ties, including friends, neighbors, past and present coworkers, friends' coworkers, etc.

The city we study is divided into 1406 cells: 917 cells in the city proper (the urban center of the city) and 589 cells in surrounding counties (see Figure 2 for a section of the city map).[12]. A cell is similar to but smaller in size than a census block in the U.S. A lower level of geographical unit is a *location*, the geographic position returned by a tower station, which represents a building or an establishment within a cell. The median and average number of distinct locations in a cell is seven and thirteen, respectively.

Constructing individuals' workplace and residence using the recorded geocode is the most crucial step of our analysis. Since we do not directly observe place of work or employment status, we take a conservative approach that mitigates the extent of measurement error in our work-related variables. We focus on individuals with work location for at least forty-five weeks, a period long enough to precisely identify job switches. This gives us 560k individuals.[13] After

[11]Phone usage during 7am-9am and 6pm-10pm is excluded as people are likely on the move during these time intervals. All individuals have a unique primary work and residential location in each week unless the location information is missing.

[12]These cells are constructed by Company A for billing purposes.

[13]The sample attrition is driven by several factors. First, China's cellphone market is dynamic with a high fraction of subscribers switching carriers in each period. In addition, the primary location information is missing for weeks when individuals travel out-of-town, experience frequent location changes (common for unemployed or part-time workers, salesman, etc.), or have limited phone usage (which is common toward the end of each month for subscribers on prepaid plans).

further restricting to individuals with at most two primary working locations throughout the sample period (which excludes sales persons and people with short-term business travels and family visits) and complete demographic information, our final sample is reduced to 456k users. We carry out the core empirical analysis using this sample and conduct robustness checks in Section 4.5 with less stringent sample selection criteria.

We identify individual i as a *job switcher* if the following couple of criteria are satisfied. First, as shown in Figure 3, a job switcher is observed in both work locations for at least four weeks each. Primary locations that are visited during the working hours on a daily (weekly) basis for a month in a row are likely to be a work location, instead of shopping centers or recreation facilities. Second, the distance between these two locations is at least 1km. We choose the cutoff of 1km to reduce the possibility of erroneously identifying someone as a switcher since individuals' work locations are geocoded up to a radius of 100-200 meters.[14] Among the 456k users in our final sample, 8% (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% (Nie and Sousa-Poza, 2017). China's job-to-job mobility is lower than Western countries (e.g., 15-18% in European Union as documented in Recchi (2009)), partly because of the Hukou system that imposes significant moving costs on individuals moving 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 new work locations that are spread in 1,100 cells, about two-thirds of which are in the city proper with the remaining ones in surrounding counties.

Vacancy Data High quality data on vacancies and the matches with workers are extremely hard to find. To gauge the dynamics of labor market conditions, we collected listings from the top two largest online job posting websites, zhilian.com and 58.com, from August 2016 to Feburary 2018.[15] These websites hold on average 10,000 job postings per month. We obtained a total of 121,055 postings and merged them to our call data based on locations.

For each posting, we observe the posting date, job title and description, full time or part time, qualification (minimum education and years of experience), monthly salary (in a range), firm address, firm size (number of total employees), and firm industry. We group these postings into eight occupations using the 2010 U.S. occupation code, based on the job title and description. Popular occupations include Professionals (26.70%), Service (26.61%), Sales and

[14] The average distance between cell centroids is 1km.

[15] Zhillian.com reported a 27.5% 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% of the market in the fourth quarter of 2017 and serving more than four million firms (`http://www.ebrun.com/20161230/208984.shtml`).

Office administration (19.24%), Management (17.47%), followed by Education, Legal, Arts and Media (11.53%), Farming, Fishing, and Construction (6.44%), Production and Transportation (2.29%), and Health related (1.45%). Industries are classified in ten sectors based on the 2012 US census codes (See Appendix A for more details).

Administrative Firm-Level Records Our vacancy postings report a wide salary range (e.g., annual salary of 25k-40k RMB). Using the mid-point of the reported salary range delivers a rather flat wage profile across industries: jobs in the construction sectors are entitled to a similar salary as jobs in professional services. It is also common to withhold salary information. Finally, a sizable fraction of workers' compensation consists of non-wage benefits, including bonuses and commissions, paid vocations, health and unemployment insurance, etc. (Cai et al., 2011).

We utilize two firm-level administrative datasets to obtain information on wages and benefits, local industry composition, and firm attributes. The first is the annual National Enterprise Income Tax Records from 2010 to 2015 that are collected by the State Administration of Taxation, which contains firm ID, industry, ownership (SOE, private, foreign, etc.), the balance sheet information (revenue, payroll, employee size, etc), as well as tax payments. This database includes most large companies (major tax payers) and a sample of small- to medium-sized firms, covering about 85-90% 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 (SAIC). Our final data set contains firm location, industry, ownership type (whether state owned or private), employee size, revenue, wage payroll, capital, for a total of between five to ten thousand firms.[16]

Most of the firms are private (85.6%), followed by state-owned (7.0%), foreign (0.7%), and other ownership types (6.6%). Over 60% of firms belong to the manufacturing sector, much higher than the national average of 32% (China's National Statistic Bureau), reflecting the industrial focus of the city. Using the average payroll as a measure of job compensation, jobs in non-manufacturing firms are paid significantly higher than those in manufacturing firms, demanding nearly a fifty-percent premium (the average annual wage being 32,005 RMB vs. 20,609 RMB).

Housing Price Our main data source does not contain individuals' socioeconomic measures like wealth or income. To overcome this data limitation, we scraped 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

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

average housing price per square meter, year built, total number of units, average size, and street address. About 64% of the cells in city proper and 20% of cells in surrounding counties can be merged with residential neighborhoods in Anjuke.com.

These data sources allow us to create a large number of attributes for each location and cell, including the most common occupation among job postings, industry composition, number of employees and vacancies, average wage, and housing price. For each individual in our final sample, we observe their work and residential location, friends, neighbors, friends' workplaces and home locations, etc.

Chinese Labor Market China's labor market has a few noticeable features. First, relative to other developing countries, China has a high female labor participation rate. Due to the employment pressure with a large population, China has instituted a mandated (early) retirement age, which is 55 for female workers and 60 for male workers.

Established in the 1950s, China's hukou system categorizes individuals as agricultural or non-agricultural people based on their place of birth and is intended to anchor peasants to the countryside. According to Zhang and Wu (2018), China's urban labor market has a two-tier system, where rural migrants in urban cities take jobs with low wages and long working hours and are often denied of social benefits. The large divide between urban cities and rural areas in terms of job opportunities, social benefits, and amenities (education, health care, etc.) has created a high fraction of migrant workers.

State owned enterprises (SOEs) account for approximately 30 to 40 percent of China's GDP and 20 percent of total employment (State Assets Supervision and Administration Commission 2017). Many of the SOEs have appeared in the Fortune Global 500 list and become the largest conglomerates in the world. Most private and foreign companies trail behind SOEs in terms of firm size and revenue. Employment opportunities at SOEs are sought after for their job security, generous benefits, and sometimes higher wages than those in non-state sectors.

Similar to U.S. and European countries, referrals are common among Chinese workers. Figure 4 compares the popularity of different job search methods among Chinese and U.S. workers, where the red and blue bars represent data from the 2014 China Family Panel Studies and the 2014 U.S. Current Population Survey, respectively. Workers in China are more likely to rely on informal search methods (38% of workers in China find jobs through friends, compared to 30% in the U.S.), while formal search methods (ads, job agencies, or contacting employers directly) are more prevalent in the U.S.. 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 landing a job.

Summary Statistics: Demographics and Social Ties Table 5a presents descriptive statistics of individuals in our sample. Thirty-six percent of users are female and ninety percent of users are younger than sixty, reflecting the higher mobile phone penetration among males and the younger population. Three quarters of our sample users are born in the local province, with the rest having migrated from other provinces. Thirty-nine percent of users are born in the city proper (the urban center of the city).

The bulk of our analysis focuses on job switchers and their social network. Individual 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.[17] As Table 5b illustrates, job switchers bear similar demographics as non-switchers, except for age. Job switchers are more likely to be in their thirties and are on average two years younger than non-switchers. They are less likely to be migrants and have a smaller fraction of friends using Company A's mobile service, although the magnitude of these differences is very modest.

The call data consist of rich information on users' social networks, but only report work locations for Company A's subscribers. On average, 50% of an individual's friends are intra-firm connections. One might be worried about potential sample selection bias if Company A's subscriber network over-represents 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 with comparable user social demographics. To test the robustness of our results with respect to potential sample selection bias (our core analysis is limited to Company A's subscribers), Section 4.2 separates individuals with friend coverage above the median from those with coverage below the median and documents very similar findings.

4 Empirical Analysis: Referral-Based worker flow

Results in Section 2 provide strong evidence that information flows are predictive of worker flows. Here we examine the importance of a specific channel: referrals, the human carrier of information on job openings and vacancies. When individuals move to a *pre-existing* friend's workplace, such a friend is defined as a referral. We limit individuals' networks to those formed prior to job switches throughout this section (except when noted otherwise). This avoids endogenous links formed post the job switch.

Among the 38,102 job switchers observed in our sample, 4,703 workers' (12%) friend location

[17] There are several possible definitions of social contacts. One alternative includes everyone who makes a phone call to *and* receives a phone call from i at lease once during our sample period. These two different definitions of social contacts lead to very similar results. Section 4.5 conducts robustness checks on the definition of social contacts.

information is missing (Panel A of Table 6). Among the switchers with non-missing location information for at least one friend, 25% find a job through a referral. Note that this should be interpreted as a lower bound as we only observe friends' locations if they are have forty-five weeks of non-missing work location information. As discussed in Section 3, forty-five weeks ensure the accuracy of identified job switches, but may under report the fraction of referred job moves. In Panel B, we relax the friend sample to all social contacts with at least four weeks of non-missing work locations. Among switchers with friend location information, 43% move to a friend. In light of this difference, Sections 4.2 to 4.4 present estimates with our preferred friend definition where friends are social contacts with at least forty-five weeks of work information, while Section 4.5 repeats these analysis using friends with at least four weeks of work information. Results are very robust to this alternative friend definition.

One might be concerned that the referral definition in our sample suffers from several confounding factors, in particular, sorting or homophily. We proceeds as follows. We first present the time series variation of information exchange between job seekers and referral vs. non-referral friends (Section 4.1). Then we exploit rich variation in our dataset and conduct a battery of tests to argue that our estimate of the referral effect is not driven by confounding factors (Section 4.2). Finally, we evaluate the benefits of referrals to workers (Section 4.3) and firms (Section 4.4).

4.1 Event Study

Our detailed calling records allow us to examine communication patterns between a job seeker i and his referral vs. non-referral friends over time. To our best knowledge, this is the first empirical analysis that directly measures information exchange between job seekers and referrals.

We first present evidence that the weekly number of contacts individuals reach out to prior to their job change is stable. Using all friends of job switchers (including inter-carrier contacts), Figure 5 illustrates that there are no spikes in the number of friends communicated with during the weeks leading to the job switch. The average number hovers between twenty three and twenty five for most weeks, with a modest decrease just prior to the switch. This provides evidence that social links established prior to job switches are likely exogenous; otherwise we should expect a spike before the job switch. The weekly number of all contacts communicated with post the job switch is somewhat higher, which is intuitive and reflects new relationships formed at the current place of work.

To examine the dynamics in information flow between referrals and referees, we regress the phone call frequency between individual i and his friends on the event window of eleven months before and after the job switch, with a rich set of fixed effects:[18]

$$\text{Freq}_{ijt} = c + \sum_{s=-11}^{10} b_s \text{Non-Referral}_{ij} D_s + \sum_{s=-11}^{10} \gamma_s \text{Referral}_{ij} D_s + \lambda_i + \tau_t + \epsilon_{ijt}$$

where Freq_{ijt} is the number of calls between caller i and his friend j in month t, D_s is the event dummy relative to the month when the job switch occurs, λ_i is individual fixed effects, and τ_t is month fixed effects. Referral_{ij} takes value one if switcher i moves to a friend j's workplace during the sample period, and zero otherwise. Figure 6 plots the regression coefficients and their confidence intervals for referral pairs (γ_s) and non-referral pairs (b_s) separately. Note that the confidence intervals are much tighter for the estimates of non-referral pairs (b_s) because the sample is much larger. There are 253k switcher-referral-month observations relative to 4.9m switcher-non-referral-month observations.

The communication patterns between referral and non-referring pairs are distinct, even after controlling for a rich set of fixed effects. First, switchers have more frequent calls with referrals than non-referrals, suggesting that referral friends are stronger ties. Second, the intensity of information flow between workers and their referrals exhibits an inverted-U shape that peaks at the time of job change. In contrast, the information flow between non-referral pairs remains stable throughout the sample period, with no noticeable change in the months leading to the job switch. Lastly, communication intensity between referrals and referees remains elevated post job switch and is noticeably larger than that between non-referral pairs. Information flow increases with the dimension of social interaction, as referrals and referees are friends before the job switch and become friends and colleagues afterwards.

One might be worried that individuals could share with friends news about their job change after obtaining one, which would also lead to intensified communication before the job switch. However, if this were true, we should expect to observe a spike in the communication volume with *both* referral *and* non-referral friends. The fact that we do not see such an increase with non-referral friends indicates that the communication between workers and referrals is unlikely to be mainly driven by workers informing friends of their job change.

4.2 Referrals and Work Location Choices

Buttressed by the evidence from the event study that friends provide job seekers with useful information on job opportunities, we turn to a regression framework to quantify the magnitude of the referral effect in shaping job seekers' location choices. Specifically, we compare the propensity for an individual to find a job at a friend's workplace with that of getting a job in

[18]Running regressions separately for referral-pairs and non-referral-pairs delivers similar patterns.

a nearby location, using the following model:

$$M_{il} = \beta Friend_{il} + \sum_{k=1}^{K} \beta_k X_{ki} + \lambda_c + \varepsilon_{il} \tag{1}$$

where the dependent variable M_{il} is one if i moves to location l. $Friend_{il}$ is a dummy variable for having at least one friend working in location l and λ_c denotes cell fixed effects that control for unobserved location attributes (number of job vacancies, industrial composition, number of locations, etc.). Demographic controls $X_{ki},\ k = 1, ...K$ contain gender, migrant status, and age group categories (age 25-34, age 35-44, age 45-59, and above 60). We also include i's total number of social contacts (irrespective of carriers) to capture differences in personality and social outreach.

Note that we only consider job switchers (people who have found a job). Analyzing how referrals affect the probability of looking for a job (the extensive margin) is interesting but outside the scope of our analysis. In addition, we restrict individual i's choices to locations *within* the cell c that contains his new workplace. This is done on purpose. Job location choices are influenced by many factors including local amenities, industry composition, local labor demand, and commuting distance, many of which cannot be directly controlled in our framework. Limiting individuals' choices to locations within the cell of his new workplace greatly reduces the extent of heterogeneity across locations and allows us to better isolate the effect of referrals from competing explanations of location choice.

The coefficient of interest is β which captures the referral effect. There are two main threats to a causal interpretation. First, a positive correlation can arise in a scenario with exogenous worker flows from one area to another. For example, firms sometimes relocate, consolidate, or open new plants at different locations. If employees are relocated in different time periods, the estimated β could capture flows of workers moving to the work location of pre-existing contacts (colleagues). We tackle this problem by adding the *interaction* of the origin and destination cell fixed effects:

$$M_{il} = \beta \text{Friend}_{il} + \sum_{k=1}^{K} \beta_k X_{ki} + \lambda_{\tilde{c},c} + \varepsilon_{il}$$

where $\lambda_{\tilde{c},c}$ is a dummy for the pair of individual i's previous ($\tilde{c}$) and current work cell (c). This is a demanding specification where the key coefficient β is estimated by the within-origin-destination variation. We essentially compare individuals with the same origin-destination cell pair but different friend network and examine their choice of locations in the same neighborhood.

The second long-standing challenge in the literature using observational data is the difficulty to distinguish 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 instead of 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 location specialized in jobs requiring similar skills.

We leverage on the richness and structure of our data and propose the following battery of tests. First, we limit our analysis to workers for whom there is at least another location within the same cell that has vacancy listings in the same occupation as the one that he takes.[19]. This mitigates the concern that individuals sort into friends' locations which 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 have 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.

Third, we compare between friends who work vs. live at location l. Sorting into friendship would imply a similar effect for these two types of friends. On the other hand, larger estimates for friends working in location l would be consistent with referrals: affiliation with the workplace may enable friends working there to have an information advantage of job openings.

Results Table 7 reports the coefficient estimates for model (1). Column 1 only controls whether there is a friend in a given work location. Column 2 adds demographic variables: gender, dummy for each age group, migrant status based on the birth county, and total number of friends prior to the job switch.[20] Columns 3 and 4 repeat the first two columns with cell fixed effects for the new workplace. Columns 5 and 6 further include 17k fixed effects for the pair of old and new work cells.

The mean propensity to choose a given location is 0.09. The coefficient for the referral effect is economically large, precisely estimated, and stable across all columns in Table 7, ranging from 0.36 to 0.38. The probability of moving to location l increases by four times with a friend working there. Adding demographic controls and interaction of origin-destination cell fixed effects have little impact on the key parameter estimate. Table 8 illustrates that having a friend at the new workplace is an extremely important predictor for individuals' job location choice. Columns 1 and 2 of Table 8 replicates the last column of Table 7, except that Column 1 excludes the friend dummy. Adding the friend variable for a sample of nearly one million observations boosts the R-squared by 2.5 times from 0.06 to 0.14. The R-squared improves

[19] 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% of all postings at the same location.

[20] Friendship that is formed post the job switch is endogenous and excluded from all regressions.

further in Column 3 that controls for the number of calls between individual i and location l prior to the job change, echoing results documented in Section 2. We next conduct a goodness-of-fit exercise similar to that performed in an independent study by Buchel et al. (2019) and report the percentage of correct predictions (the second to the last row) where the observed location choice has the highest fitted linear probability. The fraction of correct predictions is 9.3% in Column 1 and jumps to 25.0% and 28.3% in Columns 2 and 3, respectively, when Friend_{il} and number of calls to l are included sequentially.

One might be concerned about sample selection bias since information on work location is missing for friends outside Company A's subscriber network. Table 9 splits the sample based on whether the friend coverage is above or below the median (the cutoff is 48%) and replicate Columns (2), (4), and (6) in Table 7. The difference in the friend coefficient is modest and insignificant with cell-pair fixed effects (0.36 vs. 0.38, the last two columns).

To evaluate whether our finding is driven by sorting, we conduct the three tests described above in Table 10. All columns include the old and new work cell-pair fixed effects and demographic controls. Columns 1 and 2 limit to the subset of switchers with at least one alternative work location within the same cell that has openings in the same occupation as the one they take. This produces a modest impact on the estimate: the coefficient of Friend_{ij} changes from 0.37 to 0.36. Columns 3 to 6 use the same sample as that in Column 1 and 2. Columns 3 and 4 contrast friends currently in the new work location with friends who recently moved away, while Columns 5 and 6 compare friends working vs. living there. In both cases, friends currently working in the new location have a much larger impact on the choice probability: they are five times more influential than friends who recently moved away and 150% more effective than friends living in the same location. The differences in parameter estimates are statistically significant at the 1% level. These results cannot be reconciled with sorting and provide evidence that referrals at work carry useful information that facilitates the matching between workers and job openings.

Pathway The information channel can operate through different mechanisms. 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 two stories, 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 living far away from the new work location, with limited work experience, or changing industrial sectors are likely to be disadvantaged in obtaining information about new openings. Similarly, employers are less likely to be knowledgeable about these workers. In Table

11, we interact Friend_{il} with the distance between old and new work place, distance between home and new work location, dummy for young workers (between 25 and 34), moving from rural to urban, and changing sectors.[21] Referrals facilitate job transitions in *all* these situations, especially for rural workers migrating to urban areas and for people changing industry sectors. For these two groups of individuals, the point estimate of the referral effect is 0.71 and 0.55, respectively, a significant boost from the base estimate of 0.37. In Column 7 of Table 11, we interact Friend_{il} with the demeaned number of calls between individual i and location l prior to job-switch. The referral effect increases with the calling intensity: one hundred calls are associated with a two percentage point increase in the probability of moving to a friend's place, consistent with findings in Gee et al. (2017) that examines the effect of strong social ties on job search using Facebook friends.

Comparison with the Literature How do our results compare to the existing literature on referral effects? There are two common approaches of inferring referrals in observational studies. The first defines referrals as residential neighbors, pioneered by Bayer et al. (2008). Using data from the Boston metropolitan area, they define friends as individuals living in the same Census block. The other strand of literature assumes that social interactions are stronger within the ethnic group and defines friends as co-workers from the same minority group (Bandiera et al., 2009; Dustmann et al., 2016). We re-estimate model (1) using these two definitions of friendship and report the results in Table 12. 'Residential neighbor' is a dummy variable that takes value one if individual i has a neighbor who shares the same residential location as i working in location l. Ethnicity is inapplicable in China's context and is replaced with birth county as the literature has documented strong social ties among individuals from the same birth region (Zhao, 2003).[22] 'Same birth county' takes value 1 if individual i has a co-worker 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 friends while Column 4 compares coworkers sharing the same birth county with friends working in the same location.

Results in Table 12 confirm the findings in the literature that neighbors and coworkers from the same birth counties are important. The coefficients on neighbors and same birth county are 0.23 and 0.11, 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 more than doubles the probability of switching to location l. On the other hand, friends whom workers communicate with dominate either type of social ties by a large margin. The difference in magnitude is both statistically significant and economically sizable, and in the case of 'same

[21] The sample size drops in Column (6) because the dominant sector is undefined for a large number of locations whose postings from the most common sector account for fewer than 33% of all postings.

[22] In China, counties are a lower level of geographical unit and smaller than cities. There are seventeen counties in the city we study.

birth county', the effect of friends is three and half times as large.

Attributes of Referrals and Referees To examine the characteristics of workers who find a job through referrals and of friends who provide referral information, we use a dyadic regression framework where the probability that individual i moves to friend j's workplace is a function of their attributes:

$$M_{ij} = \sum_{k=1}^{K} \alpha_k X_{ki} + \sum_{k=1}^{K} \beta_k X_{kj} + \sum_{k=1}^{K} \gamma_k X_{k,ij} + \lambda_c + \varepsilon_{ij} \tag{2}$$

where X_{ki} and X_{kj} include gender, age, and birth county dummies for switcher i and friend j. $X_{k,ij}$ includes dummies for the same gender, same birth county, and an absolute difference in age. We limit the sample to all dyads between job switchers and their friend links prior to the job switch. Thus we are comparing dyad $\{i, j\}$ where i moves to j with dyad $\{i, m\}$ where i does not move to m.

We limit the regression sample to the set of 10,520 switchers who find a job at a friend's workplace. Column 1 of Table 13 includes all eligible dyads with non-missing demographic information, for a total of 93k observations. Column 2 limits to switchers for whom there are other job openings in the same cell. There are 88k observations. The dependent variable mean is 0.14. Females and migrant workers are more likely to receive referrals. There are strong assortative patterns in referral provision. Females are less likely to provide referrals on average but are more likely to provide referrals to other women. Similarly, workers are more likely to refer other workers from the same hometown county. This is consistent with recent studies documenting that in China community networks based on birth county facilitate entry and growth of private enterprises (Dai et al., 2018). Finally, older workers are more likely to provide referrals and individuals with a similar age are more likely to refer jobs to each other, though both effects are modest. Given that females and migrant workers are disadvantaged in urban labor markets (Gagnon et al., 2014; Blau and Kahn, 2017; Abramitzky and Boustan, 2017), these results provide suggestive evidence that referrals improve labor market inequality.

4.3 Referral Benefits To Workers

Our evidence thus far suggests that referrals carrying information about job openings help job seekers find new employment opportunities. In this section, we examine whether referrals improve referees' labor market outcomes conditioning on finding a job. Our framework to analyze the benefit of referrals is conceptually similar to model (1):

$$Y_{ilr} = \beta \text{Friend}_{ilr} + \sum_{k=1}^{K} \beta_k X_{ki} + \lambda_c + \alpha_r + \varepsilon_{ilr} \tag{3}$$

where Y_{ilr} denotes the labor market outcome of worker i living in residential cell r and switching to work location l in cell c. We control for the same set of demographic variables as in model (1). As we do not observe individuals' socioeconomic information such as education and wealth, we include in all regressions the residential cell fixed effect (α_r) as a proxy for one's social status (luxurious complexes vs. low-income neighborhoods).

We have constructed five different measures of job quality. Our first measure is the expected wage at the new job, measured by the average payroll (in thousand RMB) among firms in the same location weighted by their number of employees.[23]

As wage dispersion is primarily driven by across-firm instead of within-firm differences (Card et al., 2018) and that individuals' housing value is correlated with labor income, we use coworkers' housing price as a second measure to proxy for monetary compensation. Specifically, we construct the difference between average housing price of co-workers at the new workplace and that of co-workers at the previous job. Large positive differences are more likely to be associated with increases in wage and other pecuniary benefits.

The other three measures on job amenities include whether moving from a part-time job to a full-time job, changes in the commuting distance, and whether switching from a non-SOE firm to a SOE, as SOEs are sought after for their job security and pension benefits (Zhu, 2013). Though none of these measures of job outcomes is perfect, collectively they speak to both financial and non-financial aspects of job quality.

Results As our labor market outcomes are constructed from different data sources, the number of observations across specifications in Table 14 varies from 15,881 to 29,117 and reflects the varying extent of missing observations. Referral jobs pay higher expected wages than non-referral jobs. The point estimate of the wage premium is 620 RMB, about 2% of the average wage reported in our sample.[24] Turning to differences in coworkers' home values in the new vs. old workplace, referral jobs are associated with a 0.5% higher housing price per square meter, where the average housing price in the city is 13 thousand RMB ($2,000) per square meters.

Having at least one friend at the new workplace helps to increase the probability of moving from part-time to full-time jobs by one percentage point, which is a 7% increase in the likelihood

[23]For each firm in the tax data, we assign it to the nearest location in our sample and cap the distance at 500 meters. Firms that are farther away are dropped. For 79% of job switchers, the wage information is obtained from a firm within 300 meters. The employment weighted annual average payroll reflects more accurately the average worker's compensation.

[24]The annual wage is measured in thousands of yuan and the mean is 32.

of working full-time.[25] Twenty-nine percent of job switches involve a shorter commute. Referred jobs are associated with a 30% increase in the likelihood of working closer to home. Finally, having a friend at the new workplace raises the probability of moving to a SOE firm by one percentage point, which is a 9% increase from the mean (0.12).[26]

These results suggest that referrals not only facilitate the match between workers and vacancies but also improve referees' labor market outcomes through mitigating information frictions in the hiring process.

4.4 Referral Benefits To Firms

With a few exceptions, most studies on referral effects abstract away from analyzing outcomes of firms hiring referees, as comprehensive data on the performance of both employees and employers are hard to obtain.[27] We merge our calling data with administrative firm-level data based on locations and examine variation across a large number of firms in different industries.

We successfully merged between 5k and 10k firms, 67% of which are manufacturing firms that require production facilities. Our main specification focuses on locations matched to large firms with more than one hundred employees, about 20% of our sample. While this choice significantly reduces the sample size, it mitigates measurement errors as there could be multiple firms in the same location and it is difficult to match workers to firms. The average employment for large firms is 150. It is thus likely that these firms occupy an entire location. In the rest of this section, we use "location" and "firm" interchangeably. Appendix Table A2 reports results from replicating the analysis using all firms. The referral effects are stronger, with the key coefficients more significant both statistically and economically.

We compare the performance of firms hiring through referrals with that of firms not hiring through referrals via the following model:

$$Y_i = \gamma \text{Referral}_i + \sum_{k=1}^{K} \beta_k Z_{ki} + \lambda_c + \varepsilon_i \tag{4}$$

where i denotes a firm. We examine five measures of firm performance (Y_i): (1) worker inflows, (2) worker outflows, (3) net inflow of workers, (4) match rate, measured by the number of hires over vacancies, and (5) firm growth rate, measured by the number of hires over total number

[25] Hours worked is measured by the duration of phone usage during a workday at the workplace. For example, if an individual uses his phone at 10am and then at 4pm in the work location, then hours worked is 6. This is likely an under-estimate of the actual hours worked. Part-time (full time) is defined as thirty hours or fewer (more than thirty hours). On average, individuals work full-time 17% of the week, reflecting the conservativeness of our measure.

[26] A workplace is classified as 'SOE' if the majority of workers there are employed by SOE firms.

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

of employees. We use $log(Y + 1)$ for the first three outcomes to include observations with zero values and convert the coefficient estimates to semi-elasticities, the percentage change in firm outcomes when they hire through referrals. The last two measures are in logs, hence the coefficients are directly semi-elasticities. We limit to locations with at least one job switcher, otherwise the estimate of γ will be inflated artificially 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 switching to firm i has a friend working in firm i and zero if none of the hires is based on referrals, and λ_c denotes work cell fixed effect, the same as in model (3). The set $\{Z_{ki},\ k = 1, ...K\}$ denotes various firm attributes and employees' characteristics. Firm attributes include firm age, average number of employees (firm size), dummies for eighteen different industries, large firms, and SOEs, and average real capital from 2010 to 2015. We also control for the average employment growth rate from 2010 to 2015 to capture pre-existing trends. Finally, we include a firm's referral network size, which is defined as the number of unique social contacts owned by the pre-existing employees who work in firm i prior to referrals. Worker attributes include the shares of female workers and migrants, average age of employees, and average housing price of the pre-existing employees.[28]

Results The parameter estimate γ captures the effect of using referrals on firms' performance. To the extent that firms growing fast are more likely to hire through employee referrals, our estimate could be inflated. We tackle this problem by estimating model (4) with an increasing set of variables that help to control for firms' growth rate and employee quality. We report the results in Tables 15 and 16.

The Referral$_i$ coefficient estimates are remarkably similar across different sets of controls for firm and worker attributes, indicating our results are unlikely to be inflated by unobserved firm or worker characteristics. Firms hiring through referrals are associated with more hires, better matching rates, and a higher growth rate. Using referrals increases a firm's net labor inflow by 44%, enhances the job matching rate by 60% (the mean matching rate for large firms is 0.38), and raises the firm growth rate by 55% (the median growth rate is 4% for large firms). Interestingly, the effect on the number of workers leaving is insignificant despite more hiring, suggesting that friends might help to retain workers. Although our analysis in this section is largely descriptive, the fact that the estimates are robust to a detailed set of firm and worker controls raises our confidence in those estimates that they are not simply picking up unobserved characteristics related to firm and employees quality.

[28]Similar to all previous models, the number of Company A's users at each location is included in all regressions.

4.5 Alternative definition of friends

We conclude our analysis with a few additional robustness checks. Our core analysis thus far limits to friends with at least forty-five weeks of non-missing location information and observed in their primary workplace for at least twenty-three weeks. This mitigates measurement errors in job locations but omits a large fraction of friends for whom we observe fewer than forty-five weeks of location information. The downside of this restriction is that, the estimate of the key regressor, 'Friend$_{il}$', whether individual i has a friend in location j, is biased toward zero. In this section, we examine the robustness of our results to alternative sample selection criteria.

Appendix Tables A3 replicates Table 7 while including all friends who are observed in their primary location for at least four weeks. This enlarges the individual-friend pairs from 401,437 to 979,595. The estimated referral effect remains unchanged: having a friend increases the probability of moving there by 37 percentage points.

Using this alternative definition of friend, referral jobs are associated with 1.3% increase in wage premium, 0.6% increase in job-related benefits (as proxied by coworkers' housing price), a 12% increase in the likelihood of working full-time (Appendix Table A4). These effects are similar to those in our base specification. The effect on the likelihood of a shorter commute and transitioning to premium jobs is nearly identical to the base specification.

Turning to the benefit of referrals to firms, the alternative definition of friend produces slightly more pronounced results than those reported in the base specification (Appendix Table A5), suggesting referrals could greatly benefit firms in terms of hiring and future growth.

Using other and more stringent selection criteria (three months, six months, etc.) deliver similar patterns. This is comforting and suggests that different types of measurement errors associated with varying degree of stringency seem to cancel each other out.

Finally, Appendix Table A6 repeats Table 7 but defines individual i's friend as someone who both places and receives at least one call from individual i (a two-way social contact) and is observed for at least four weeks in his/her primary work location. The referral estimate is again very similar to our base specification.

5 Conclusion

This paper uses novel geocoded mobile phone records to study information exchange between spatial and social groups and examine how job information passed on from social contacts to job seekers mitigates information asymmetry and improves the labor market performance.

Our study demonstrates how newly emerging data from Global-Positioning-System (GPS) enabled devices, such as mobile phones, may facilitate our understanding of how markets function. It provides three broad lessons for future research. First, panel data at an extremely detailed level of spatial and temporal disaggregation hold great potential for overcoming traditional challenges of causal inference with observational data. For example, in our context, the possibility to identify different types of social contacts in fine geographical areas at overlapping periods of time helps to tackle one of the most difficult challenges in empirical studies: sorting. Second, big data from non-conventional sources are complementary to traditional data sources where socioeconomic outcomes are recorded. In our analysis, merging in tax records and business registry in particular has proved crucial to shed some light on how referrals benefit firms, a topic that is understudied in the existing literature. Third, information exchange is crucial to predict job move, and more importantly, social and socioeconomic diversity in communication ties matter the most. Our results suggest that interventions aiming at reducing information asymmetries and the cost of information acquisition among firms and workers would be extremely valuable to increase market efficiency.

References

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

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

Akerlof, G. A. (1970, 08). The Market for "Lemons": Quality Uncertainty and the Market Mechanism*. *The Quarterly Journal of Economics 84*(3), 488–500.

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

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

Ashraf, Q. and O. Galor (2011, December). Cultural diversity, geographical isolation, and the origin of the wealth of nations. Working Paper 17640, National Bureau of Economic Research.

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

Athey, S., D. Blei, R. Donnelly, F. Ruiz, and T. Schmidt (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., R. Cao, T. Kuchler, and J. Stroebel (2018). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy 126*(6), 2224–2276.

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

Bayer, P., L. R. Stephen, and G. Topa (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.

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

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

Blumenstock, J. E. (2018). Estimating economic characteristics with phone data. *AEA Papers and Proceedings 108*, 72–76.

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

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

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

Cai, H., H. Fang, and L. C. Xu (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., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), S13–S70.

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

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

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

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

Dustmann, C., A. Glitz, and U. Schonberg (2016). Referral-based job search networks. *Review of Economics Studies 83*, 514–546.

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

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

Fruchterman, T. M. J. and E. M. Reingold (1991, November). Graph drawing by force-directed placement. *Software: Practice and Experience 21*(11), 1129–1164.

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

Gee, L., J. Jones, and M. Burke (2017). 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.

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

Glaeser, E. L., S. D. Kominers, M. Luca, and N. Naik (2015, December). 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., C. A. Hidalgo, and A.-L. Barabasi (2008, June). Understanding individual human mobility patterns. *Nature 453*, 779.

Hastings, J. S. and J. M. Weinstein (2008, 11). Information, School Choice, and Academic Achievement: Evidence from Two Experiments*. *The Quarterly Journal of Economics 123*(4), 1373–1414.

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

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

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

Ioannides, Y. M. and L. D. Loury (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 P. Leslie (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 O. Skans (2014). When strong ties are strong: Networks and youth labour market entry. *Review of Economic Studies 82*(3), 1164–1200.

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

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*(2), 175–208.

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

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

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

Ottaviano, G. I. and G. Peri (2006, January). 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.

Rothschild, M. and J. Stiglitz (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., A. Weber, and M. A. Weynandt (2018). Coworkers, networks and job search outcomes. *ILR Review*. Forthcoming.

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

Schumutte, I. (2016). How do social networkds affect labor markets? *IZA World of Labor October*, 1–10.

Smith, V. K. and F. R. Johnson (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.

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

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

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance 16*(1), 8–37.

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

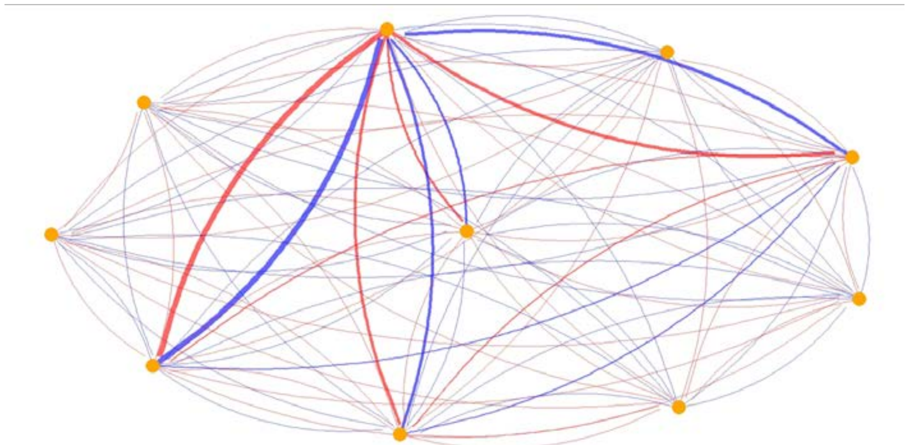
Zhang, J. and J. Wu (2018). The chinese labor market, 2000–2016. *IZA World of Labor* (437).

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.

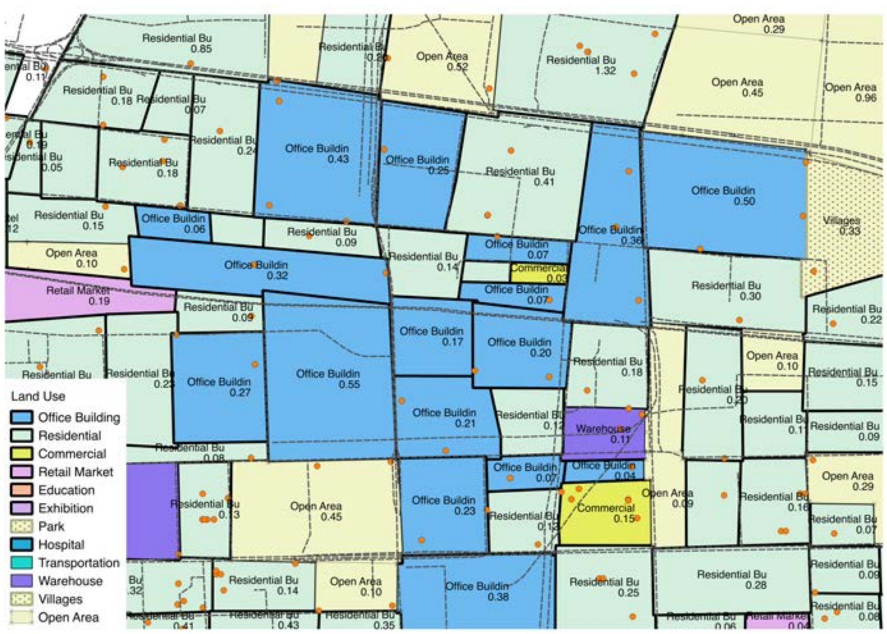
Figures and Tables

Figure 1: Information Flow and Worker Flow Among Administrative Districts



Notes: Each node is an administrative district in the city. We plot randomly selected 10 nodes out of 23. Blue (non-directional) lines correspond to the number of job switches among the pairs of nodes, with the width of each edge scaled proportionately to the number of switches. Red lines denote the average number of weekly calls, with a scaled edge-width as well. The graph is produced using the Fruchterman & Reingold algorithm that aims to distribute vertices evenly. Source: Mobile Communication Data.

Figure 2: Map of Locations and Cells in the City



Source: Mobile Communication Data.

Figure 3: Job Switch Timeline

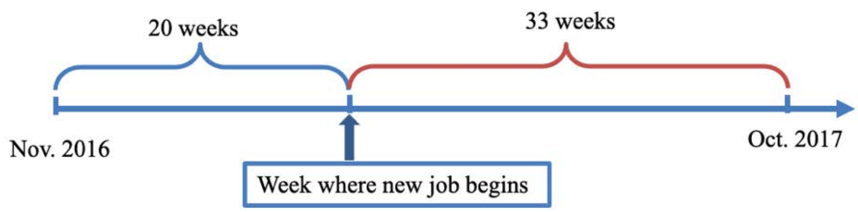
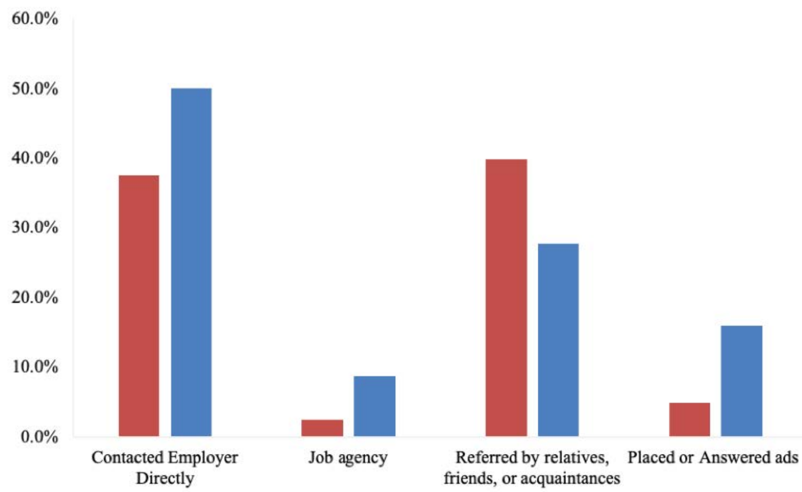
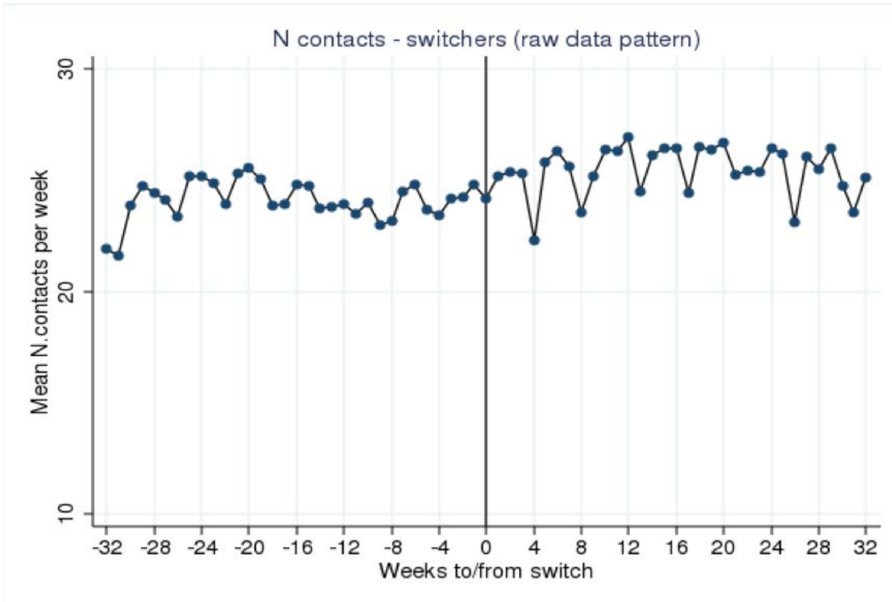


Figure 4: Job Search Methods in China vs. the U.S. (2014)



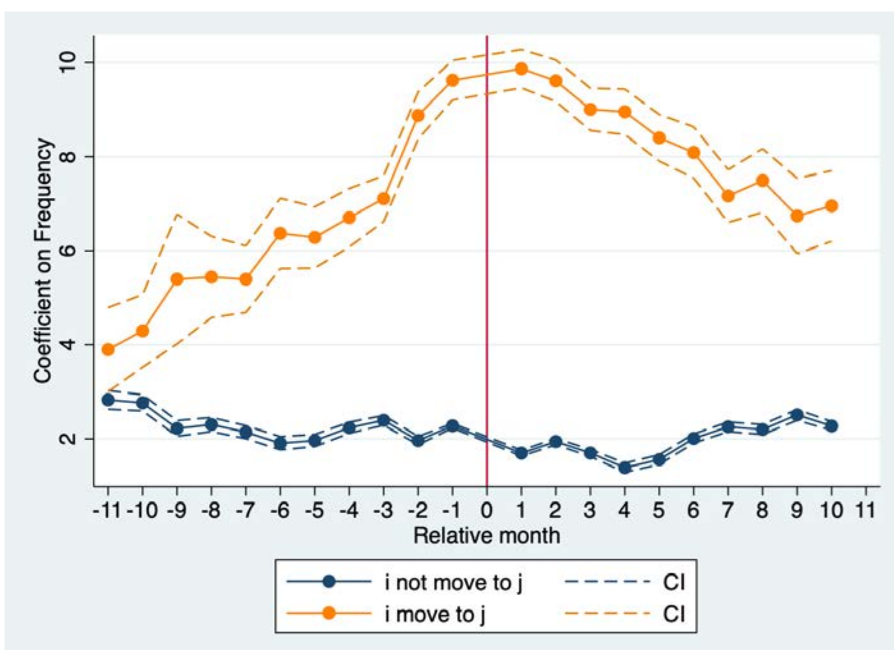
Notes: Horizontal axis reports different job search methods. Vertical axis reports fraction of each method used among job seekers. Red (blue) bars represent China (U.S.). Source: China Family Panel Studies (2014) and U.S. Current Population Survey (2014).

Figure 5: Number of Social Contacts per Week: Job switchers



Notes: The figure plots the average number of social contacts (regardless of carriers) per week who communicated with a switcher. $T = 0$ indicates the week of job switch. There are 37,099,345 switcher-contact-week observations. Source: Mobile Communication Data.

Figure 6: Event Study: Number of Calls to Referrals vs. Non-referrals



Notes: Top line represents number of calls between switchers and the referrals (Obs = 252,852). Bottom line represents number of calls between switchers and non-referrals (Obs = 4,915,656). Switcher fixed effects and month fixed effects are included in the regression. Source: Mobile Communication Data.

Table 1: Information Flows and Worker Flows

(a) At the Administrative District Level

Dep. Var.: Worker flow (i,j)	All calls	Calls from/to job switchers ***before*** switch		
		No exclusion	Excluding friends in last month	Excluding friends in last 3 month
	(1)	(2)	(3)	(4)
Information flow (i,j)	0.003***	0.09***	0.10***	0.13***
	(6.20e-05)	(0.001)	(0.001)	(0.001)
Obs.	253	253	253	253
R-squared	0.90	0.97	0.97	0.97
District i + District j fixed effects	Yes	Yes	Yes	Yes

(b) At the Location Level

Dep. Var.: Worker flow (i,j)	All calls	Calls from/to job switchers ***before*** switch		
		No exclusion	Excluding friends in last month	Excluding friends in last 3 month
	(1)	(2)	(3)	(4)
Information flow (i,j)	5.30e-05***	0.0006***	0.0006***	0.0007***
	(2.06e-08)	(1.70e-07)	(3.10e-07)	(3.97e-07)
Observations	159,856,140	159,856,140	159,856,140	159,856,140
R-squared	0.04	0.07	0.03	0.02
Location i + Location j fixed effects	Yes	Yes	Yes	Yes

Notes: In Panel (a), one unit of observation is a pair of administrative districts (i,j). There are 23 administrative units in the city. In Panel (b) one unit of observation is a pair of locations (i,j). There are 17,881 locations in the city. Dependent variable, "Worker flow (i,j)", is the total number of workers moving between area i and area j. In Column 1, "Information flow (i,j)" is the total number of calls between area i and j among all individuals. In column (2) to (4), it is the total number of calls between switchers and their pre-existing contacts.

Table 2: Out-of-Sample Prediction for Worker Flows at the Neighborhood Level: Correlations

	Dep var. Actual worker flow (i, j)					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted (linear regression + FEs)	1.02***	1.02***				
	(0.18)	(0.002)				
Predicted (linear spline)			0.97***	0.97***		
			(0.16)	(0.001)		
Predicted (Cubic Spline)					1.02***	1.03***
					(0.17)	(0.002)
Constant	0.01***		0.01***		0.004**	
	(0.001)		(0.001)		(0.002)	
Observations	987,713	987,713	987,713	987,713	987,713	987,713
R-squared	0.30	0.31	0.31	0.32	0.31	0.31
N. Knots			5	5	6	6
Neighborhood i + Neighborhood j FE	No	Yes	No	Yes	No	Yes

Notes: One unit of observation is pair of neighborhood cells. The training data consists of switches in the first six months. The prediction is based on switches in the second six months. The table reports the correlation between the actual worker flow between cell i and j and the worker flow predicted from three models. Columns (1) and (2) are a linear regression using the number of calls by switchers prior to job change as a key regressor and neighborhood fixed effects. Columns (3) and (4) fit linear splines of calls with 5 knots. Columns (5) and (6) fits cubic splines of calls with 6 knots.

Table 3: Information Diversity and Worker Flows

Dep var: Log of inflow	(1)	(2)	(3)	(4)
Social entropy	0.82**			0.95**
	(0.36)			(0.41)
Spatial entropy		-0.19		-0.58
		(0.32)		(0.36)
Income entropy			0.81***	0.70***
			(0.24)	(0.23)
Total call volume (x1000)	0.001***	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	6,161	6,161	6,161	6,161
R-squared	0.64	0.64	0.64	0.64
Cell FE	Yes	Yes	Yes	Yes
N. of Cell FE	1183	1183	1183	1183

Notes: One unit of observation is a location with at least 5 working population and 5 residential population. "Log inflow" is the log of number of people moving to a given location. Social entropy, spatial entropy, and income entropy are normalized Shannon entropy as detailed in text. Total call volume is the total number of calls (in thousand) from or to a given location. Number of carrier A users in each location is controlled in all specifications.

Table 4: Information Diversity and Worker Flows: Working vs. Residential Population

Dep var: Log of inflow	(1)	(2)	(3)
Working population's			
Social entropy	0.84**		
	(0.37)		
Spatial entropy		-0.11	
		(0.32)	
Income entropy			0.75***
			(0.23)
Residential population's			
Social entrophy	-0.10		
	(0.28)		
Spatial entrophy		-0.32	
		(0.29)	
Wealth entrophy			0.27
			(0.18)
Total call volume (x1000)	0.001***	0.001***	0.001***
	(0.00)	(0.00)	(0.00)
Observations	6,161	6,161	6,161
R-squared	0.64	0.64	0.64
Cell FE	Yes	Yes	Yes
N. of Cell FE	1183	1183	1183

Notes: One unit of observation is a location with at least 5 working population and 5 residential population. "Log inflow" is the log of number of people moving to a given location. Social entropy, spatial entropy, and income entropy are normalized Shannon entropy as detailed in text and constructed separately for the working vs. residential population. Total call volume is the total number of calls (in thousand) from or to a given location. Number of carrier A users in each location is controlled in all specifications.

Table 5: Summary Statistics

(a) All users

	Mean	SD	N
Female	0.36	0.48	435098
Age 25-34	0.29	0.46	455572
Age 35-44	0.26	0.44	455572
Age 45-59	0.27	0.45	455572
Age above 60	0.11	0.32	455572
Age (midpoint)	40.18	11.97	435194
Born in local province	0.75	0.43	455572
Born in local city proper	0.39	0.49	455572
Frac of social contacts in A	0.50	0.19	455572
Job switch	0.08	0.28	455572

(b) Switchers vs. Non-switchers

	Non-switchers			Switchers				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.36	0.48	398742	0.36	0.48	36356	-0.00	-0.45
Age (midpoint)	40.36	12.00	398817	38.23	11.49	36377	2.13***	32.49
Born in local province	0.75	0.43	417470	0.74	0.44	38102	0.01***	3.62
Born in local city proper	0.39	0.49	417470	0.38	0.49	38102	0.00	0.70
Frac of social contacts in A	0.50	0.19	417470	0.51	0.19	38102	-0.00	-0.53

Notes: The sample is restricted to workers with valid work information for at least 45 weeks during sample periods. Number of users = 455,572. 'Age' uses the midpoint of each age range. 'Frac of social contacts in A' is fraction of individuals' contacts who use company A as their service provider. 'Job switch' is a dummy for job switchers, who are identified based on the criteria described in the text.

Table 6: Job Switchers: switching to a friend's workplace

Panel A: one-way contact, at least 45 weeks	Percent	N. Indviduals.	N. dyads
Switching to a friend	0.22	8,518	135,866
Switching to somewhere else	0.65	24,881	265,571
Missing all friends' locations	0.12	4,703	
All job switchers		38,102	

Panel B: one-way contact, at least 4 weeks	Percent	N. Indviduals.	N. dyads
Switching to a friend	0.40	15,374	487,678
Switching to somewhere else	0.54	20,417	487,126
Missing all friends' locations	0.06	2,311	
All job switchers		38,102	

Notes: The sample is restricted to job switchers, identified based on the following criteria: First, a job switcher is observed in both work locations for at least four weeks each. Second, the distance between these two locations is at least 1km (the average distance between cell centroids). In panel A, a friend is a one-way contact with primary location for at least 45 weeks. In panel B, a friend is a one-way contact with primary location for at least 4 weeks. "Switching to a friend" takes value one if a switcher moves to a pre-existing friend's workplace. "Missing all friends' locations" reports the number of switchers with no valid information for any pre-existing friend. "N. dyads" reports the number of switcher-friend pairs where friends only include social contacts existed prior to the job switch.

Table 7: The Referral Effect

Dep. var. Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.38***	0.38***	0.36***	0.36***	0.37***	0.37***
	(0.004)	(0.004)	(0.019)	(0.019)	(0.016)	(0.015)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,019,706	993,030	1,019,706	993,030	1,019,706	993,030
R-squared	0.09	0.08	0.14	0.14	0.14	0.14
New work Cell FE	No	No	Yes	Yes	No	No
Old x New Cell FE	No	No	No	No	Yes	Yes
N. of Cell FE	NA	NA	1070	1067	17334	17010

Notes: One unit of observation is a switcher-location pair. A friend is a one-way contact, working at a primary location for at least 45 weeks. "Friend" is a dummy variable, which equals one if there is at least one friend working at a given location. Controls include gender, age, migrant, and network size measured by the number of social contacts (irrespective of carriers) before job switch.

Table 8: The Referral Effect: Goodness of Fit

Dep. var. Probability i switches to location j	(1)	(2)	(3)
Friend		0.37***	0.33***
		(0.015)	(0.014)
N.calls to j (x1000)			0.30***
			(0.028)
Controls	Yes	Yes	Yes
Observations	993,030	993,030	993,030
R-squared	0.06	0.14	0.14
Old x New Cell FE	Yes	Yes	Yes
N. of Cell FE	17010	17010	17010
Correct Predictions	9.28%	24.96%	28.30%
Percent increase wrt previous column		169%	13%

Notes: This table replicates Column 6 of Table 7 using model (1), except Column 1 excludes "Friend" dummy and Column 3 add "N.calls to j", the number of calls (in thousand) between switcher i and location j prior to the job switch. Correct prediction is the case where the chosen location has the highest fitted linear probability.

Table 9: The Referral Effect: by Friend Coverage

Dep. var. Probability i switches to location j	(1) Above	(2) Below	(3) Above	(4) Below	(5) Above	(6) Below
Friend	0.37***	0.39***	0.35***	0.37***	0.36***	0.38***
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Observations	524,196	468,834	524,196	468,834	524,196	468,834
R-squared	0.10	0.07	0.15	0.12	0.15	0.12
Controls	Yes	Yes	Yes	Yes	Yes	Yes
New work Cell FE	No	No	Yes	Yes	No	No
Old x New Cell FE	No	No	No	No	Yes	Yes
N. of Cell FE	NA	NA	1001	993	9659	9389

Notes: This table replicates Columns (2), (4), and (6) of Table 7. Odd columns use job switchers whose fraction of social contacts in carrier A exceeds the median cutoff, or 48%. Even columns use job switchers whose fraction of social contacts in carrier A is below the median cutoff.

Table 10: The Referral Effect: Falsification Tests

Dep. var.	Job switchers with similar job opportunities nearby					
Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.36***	0.36***	0.36***	0.36***		
	(0.017)	(0.016)	(0.017)	(0.016)		
Friend left before the job switch			0.07***	0.07***		
			(0.019)	(0.019)		
Friend working					0.31***	0.31***
					(0.015)	(0.015)
Friend living, not working					0.20***	0.20***
					(0.015)	(0.014)
Controls	No	Yes	No	Yes	No	Yes
Observations	976,923	950,938	976,923	950,938	976,923	950,938
R-squared	0.12	0.12	0.12	0.12	0.13	0.13
Old x New Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
N. of Cell FE	15788	15486	15788	15486	15788	15486

Notes: This table uses the same specifications as in Table 7, except it limits to job switchers facing at least one vacancy in the same occupation that are posted in alternative locations in the same cell.

Table 11: The Referral Effect: Heterogeneity

Dep. var. Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Friend	0.37***	0.35***	0.34***	0.36***	0.36***	0.27***	0.37***
	(0.015)	(0.016)	(0.017)	(0.016)	(0.015)	(0.015)	(0.015)
Friend*Dist(job1,job2)		0.002***					
		(0.0003)					
Friend*Dist(home, job2)			0.004***				
			(0.0007)				
Friend*Young (Age 25-34)				0.04***			
				(0.01)			
Friend* Rural to urban					0.35***		
					(0.04)		
Friend*Change sector						0.28***	
						(0.02)	
Friend*Call intensity							0.0002***
							(5.05e-05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	993,030	993,030	922,711	993,030	993,030	236,907	993,030
R-squared	0.14	0.14	0.13	0.14	0.14	0.15	0.14
Old x New Cell FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Cell FE	17010	17010	16061	17010	17010	5276	17010

Notes: this table uses the same specification as that in Column 6 of Table 7 and interacts "Friend" dummy with various measures. The median distance between old and new job is 3.8km. Median distance between new job and residential home is 3.1km. "Rural to urban" indicates switchers who move from outside the city proper into the city proper and 6 percent switchers are moving from rural to urban. Change sector is one if the switcher changes sector. Call intensity is the demeaned number of calls between switcher i and people working at location j prior to job switch. In Columns 2-6, we also control for the baseline level of the interaction.

Table 12: The Referral Effect: Comparison with the Literature

Dep. var. Probability i switches to location j	(1)	(2)	(3)	(4)
Friend Definition				
Residential Neighbor	0.23***		0.20***	
	(0.012)		(0.011)	
Same Birth County		0.11***		0.10***
		(0.004)		(0.004)
Friend, not Neighbor			0.27***	
			(0.0117)	
Friend, not Same Birth County				0.34***
				(0.017)
Controls	Yes	Yes	Yes	Yes
Observations	993,030	993,030	993,030	993,030
R-squared	0.17	0.13	0.20	0.17
Old x New Cell FE	Yes	Yes	Yes	Yes
N. of Cell FE	17010	17010	17010	17010

Notes: This table uses the same specification as that in Column 6 of Table 7. "Residential Neighbor" is a dummy which equals one if there is at least one neighbor sharing the same residential location as the job switcher in the new work location. "Same Birth County" is a dummy which equals one if there is at least one individual from the same hometown county as switcher i who works at the new location.

Table 13: Dyadic Regression: Attributes of Referrals and Referees

Dep. var Probability A switches to B	(1)	(2)
Female A	0.01**	0.01**
	(0.01)	(0.01)
Female B	-0.00	-0.00
	(0.00)	(0.00)
Both female	0.03***	0.03***
	(0.01)	(0.01)
Age A	0.00	0.00
	(0.00)	(0.00)
Age B	0.001***	0.001***
	(0.00)	(0.00)
|Age A - Age B|	-0.001***	-0.001***
	(0.00)	(0.00)
Migrant A	0.01**	0.01*
	(0.01)	(0.01)
Migrant B	-0.00	-0.00
	(0.00)	(0.00)
Both migrants with the same birth county	0.03***	0.03***
	(0.01)	(0.01)
Observations	93,196	88,207
R-squared	0.10	0.09
B work Cell FE	Yes	Yes
N. of Cell FE	1176	941

Notes: One unit of observation is a switcher-friend pair. The sample restricts to switchers who eventually switch to some friend and their friends with valid primary location information for 45 weeks. A denotes the referred person and B denotes the referral. Column (2) restricts to switchers facing at least one vacancy in the same occupation that are posted in alternative locations in the same cell.

Table 14: Referral Benefits to Workers

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dep var.	Wage at new job	ΔCoworker HP	PT to FT	Closer to Home	Non-SOE to SOE
Friend	0.62***	0.07*	0.01**	0.09***	0.01***
	(0.22)	(0.04)	(0.01)	(0.01)	(0.00)
Observations	17,615	23,323	19,431	29,117	15,881
R-squared	0.79	0.53	0.11	0.12	0.56
Residence Cell FE	Yes	Yes	Yes	Yes	Yes
New work Cell FE	Yes	Yes	Yes	Yes	Yes

Notes: Sample includes all job switchers. Same demographic controls as in Table 7. "Wage at new job" is the average annual payroll per worker in thousand RMB weighted by employee size among firms in the new work location. "ΔCoworker HP" is the coworkers' average house price (thousand RMB) in new workplace minus in old workplace. "PT to FT" is a dummy that equals one if the switcher works part-time (less than 30 hours per week) before the switch and full-time (more than 30 hours) after the switch. "Closer to Home" is a dummy that 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 (the majority employees working in SOE firms), while the previous job is not.

Table 15: Referral Benefits to Large Firms with Positive Hirings

Dep var. = Log of Inflow	(1)	(2)	(3)	(4)
Referral	0.38***	0.40***	0.40***	0.39***
	(0.07)	(0.08)	(0.08)	(0.08)
	[0.43]	[0.44]	[0.44]	[0.44]
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.78	0.79	0.79	0.79
N. of Cell FE	271	271	271	271
Outflow	(5)	(6)	(7)	(8)
Referral	0.14	0.13	0.13	0.13
	(0.09)	(0.09)	(0.09)	(0.09)
	[0.16]	[0.15]	[0.15]	[0.16]
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.76	0.77	0.77	0.78
N. of Cell FE	271	271	271	271
Net Inflow	(9)	(10)	(11)	(12)
Referral	0.37***	0.37**	0.37**	0.35**
	(0.14)	(0.15)	(0.15)	(0.15)
	[0.50]	[0.50]	[0.50]	[0.48]
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.55	0.57	0.57	0.57
N. of Cell FE	239	239	239	239
Controls				
Firm Attributes	No	Yes	Yes	Yes
Previous Firm Growth Rate	No	No	Yes	Yes
Employee Attributes	No	No	No	Yes
Cell FE	Yes	Yes	Yes	Yes

Notes: One unit of observation is a location with at least one matched firm larger than 100 employees and positive hirings. "Referral" takes value 1 if there is at least one switcher moving to a friend in the firm and 0 otherwise. Firm network size is controlled in every column, measured by the number of distinct contacts of the firm's pre-existing employees. Outcome variables inflow, outflow, and net inflow are expressed in the form of log(Y+1) to include observations with zero values. The first row reports the point estimate, the second row in round brackets reports standard errors, while the third row in square brackets reports "semi-elasticities" evaluated at the dependent variable mean. "Inflow" and "Outflow" refer to the number of switchers moving in and out. "Net inflow" is inflow minus outflow. Firm attributes include age, employee size, SOE or not, average real capital from 2010 to 2015. Previous firm growth rate is the employee growth from 2010 to 2015. Employee attributes includes share of female, share of migrants, average age of pre-existing employees. Number of carrier A users at each location is controlled in all columns.

Table 16: Referral Benefits to Large Firms with Positive Hirings – Continued

Dep var. = Log of Matching rate = hirings/N. vacancies	(1)	(2)	(3)	(4)
Referral	0.62**	0.62**	0.62**	0.60**
	(0.269)	(0.274)	(0.273)	(0.281)
Observations	[400,1000]	[400,1000]	[400,1000]	[400,1000]
R-squared	0.84	0.86	0.86	0.86
N. of Cell FE	190	190	190	190
Growth rate= hirings/firm size	(5)	(6)	(7)	(8)
Referral	0.586***	0.562***	0.558***	0.546***
	(0.107)	(0.096)	(0.097)	(0.096)
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.76	0.84	0.84	0.84
N. of Cell FE	271	271	271	271
Controls				
Firm Attributes	No	Yes	Yes	Yes
Previous Firm Growth Rate	No	No	Yes	Yes
Employee Attributes	No	No	No	Yes
Cell FE	Yes	Yes	Yes	Yes

Notes: One unit of observation is a location with at least one matched firm larger than 100 employees and positive hirings. Same controls as in Table 15. Outcome variables are in logs. “Matching rate” is measured as inflow over the number of vacancies. “Growth” is measured as inflow over employee size.

A Appendices

Occupancy Description We use job description and job title from each vacancy posting and the US 2010 occupation codes to classify the occupation for each posting. Here are the occupations that we use:

1. Management – including customer service manager, warehouse manager, production manager, hospital manager, human resource manager, CEO, retail shop manager and vice manager, sales manager, education administrator, etc.

2. Professionals – including business operation, finance operation, computer and science, social science and non-training professionals; business related, including wholesale trader, market research analyst, meeting and event planner, cost estimator, risk control worker, customer relation, accountants and auditors; computer and science related, including software developers, computer support specialists, database administrator, web developer, network and computer systems administrators, architects, biomedical engineers, mining and geological engineers, mapping technicians, nutritionists.

3. Education, legal, arts, design, and media – Education includes training professionals, preschool and kindergarten teachers, afterschool class teachers, teaching assistants, vocational training instructors, driving coach; Legal includes lawyer and paralegals; Arts, design, and media include director, model, hosts, actors, writers, photographers, video editors, news reporters, designers, magazine editors, webpage editors.

4. Service – including cashier, customer service, front desk, fire fighter, nail polisher, cleaner, massage, flight attendants, food server, cooks, laundry workers, counter attendants, security guards, surveillance control workers.

5. Sales and office administration – Sales includes retail salesperson, insurance salesperson, real estate sales agents, pharmaceutical sales representatives; Office administration includes office secretary, file clerks, curriculum consultants (in private education organizations).

6. Health related – including therapists, nurses, pharmacists, rehabilitation doctors, surgeons

7. Production and transportation – Production includes printing press operators, layout workers, general production workers, painting workers, cutting workers; Transportation includes sailors, cargo shipping drivers, drivers in general, warehouse workers, and material moving workers.

8. Farming, fishing, and construction – includes related natural resource, installation, maintenance, repair, welder, installation workers, computer repairers, maintainance workers, gardners, agricultural workers, forest workers, breeding workers, livestock cultivators.

We combine the three smallest categories (Health related, Production and transportion and Fishing etc) into 'other category' in our empirical analysis.

Table A1: Summary Statistics: Diversity Measures

Variable	Mean	SD	Median	Min	Max	N. Locations
Social entropy (working population)	0.67	0.03	0.67	0.40	0.83	6161
Social entropy (residential population)	0.67	0.05	0.67	0	0.95	6161
Spatial entropy (working population)	0.71	0.04	0.70	0.40	0.94	6161
Spatial entropy (residential population)	0.72	0.05	0.72	0	1.00	6161
Income entropy (working population)	0.46	0.11	0.46	0	0.83	6161
Income entropy (residential population)	0.46	0.10	0.46	0	0.92	6161

Notes: Entropy measures are the average normalized shannon entropy across individuals at a given location.

Table A2: Referral Benefits to All Firms with Positive Hirings

	(1)	(2)	(3)	(4)	(5)
Dep. Var.	Inflow	Outflow	Net inflow	Matching rate	Growth
Referral	0.38***	0.07***	0.45***	0.26***	0.47***
	[0.56]	[0.10]	[0.45]	[0.26]	[0.47]
	(0.03)	(0.03)	(0.05)	(0.11)	(0.06)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	[3000,5000]	[3000,5000]	[3000,5000]	[2000,5000]	[3000,5000]
R-squared	0.70	0.71	0.45	0.78	0.70
Cell FE	Yes	Yes	Yes	Yes	Yes
N. of Cell FE	707	707	651	526	707

Notes: One unit of observation is a location with at least one matched firm and positive hirings. Same controls as in Table 15.

Table A3: The Referral Effect: Robustness

Dep. var.							Similar vacancies	
Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Friend	0.37***	0.37***	0.35***	0.35***	0.37***	0.37***	0.36***	0.36***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,019,706	993,030	1,019,706	993,030	1,019,706	993,030	976,923	950,938
R-squared	0.16	0.16	0.20	0.20	0.20	0.20	0.184	0.184
New work Cell FE	No	No	Yes	Yes	No	No	No	No
Old x New Cell FE	No	No	No	No	Yes	Yes	Yes	Yes
N. of Cell FE	NA	NA	1070	1067	17334	17010	15788	15486

Notes: Same specification as in Table 7. A friend is a one-way contact who works at his/her primary location for at least four weeks.

Table A4: Referral Benefits to Workers: Robustness

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dep var.	Wage at new job	ΔCoworker HP	PT to FT	Closer to Home	Non-SOE to SOE
Friend	0.40**	0.08**	0.02***	0.09***	0.01**
	(0.19)	(0.03)	(0.01)	(0.01)	(0.00)
Observations	18,595	24,835	21,016	31,013	16,789
R-squared	0.79	0.52	0.10	0.12	0.56
Residence Cell FE	Yes	Yes	Yes	Yes	Yes
New Work Cell FE	Yes	Yes	Yes	Yes	Yes

Notes: Same specification as in Table 14. A friend is an one-way contact who works at his/her primary location for at least four weeks.

Table A5: Referral Benefits to Large Firms: Robustness

	Large firms with positive hirings				
	(1)	(2)	(3)	(4)	(5)
Dep. Var.	Inflow	Outflow	Net inflow	Matching rate	Growth
Referral	0.49***	0.28***	0.66***	0.73***	0.62***
	[0.55]	[0.33]	[0.90]	[0.73]	[0.62]
	(0.07)	(0.08)	(0.11)	(0.27)	(0.09)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	[600,1000]	[600,1000]	[600,1000]	[400,1000]	[600,1000]
R-squared	0.80	0.79	0.62	0.87	0.85
Cell FE	Yes	Yes	Yes	Yes	Yes
N. of Cell FE	271	271	239	190	271

Notes: Same specification as in Table 15. Friends include one-way contacts who work at their primary location for at least four weeks.

Table A6: The Referral Effect: Robustness Two

Dep. var.							Similar vacancies	
Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
havefrd_2wayL	0.37***	0.37***	0.35***	0.35***	0.36***	0.36***	0.36***	0.36***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,019,706	993,030	1,019,706	993,030	1,019,706	993,030	976,923	950,938
R-squared	0.04	0.04	0.10	0.10	0.09	0.09	0.08	0.08
New work Cell FE	No	No	Yes	Yes	No	No	No	
Old x New Cell FE	No	No	No	No	Yes	Yes	Yes	Yes
N. of Cell FE	NA	NA	1070	1067	17334	17010	15788	15486

Notes: Same specification as in Table 7. A friend is a two-way contact who works at his/her primary location for at least four weeks.