Information, Mobile Communication, and Referral Effects

Panle Jia Barwick* Yanyan Liu† Eleonora Patacchini‡ Qi Wu§

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 its role in urban labor markets. We present the first empirical evidence that information flow (measured by call volume) correlates strongly with worker flows, a pattern that persists at different levels of geographic aggregation. Conditional on information flow, socioeconomic diversity in the source of information (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 social contacts 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 sectors. Firms receiving referrals are more likely to have successful recruits and experience faster growth.

Key words Information, Mobile Communication, Urban Labor Market, Social Networks, Entropy

JEL Classification: R23, J60, L15

*Department of Economics, Cornell University and NBER. Email: panle.barwick@cornell.edu

†International Food Policy Research Institute. Email: Y.Liu@cgiar.org

‡Department of Economics, Cornell University and CEPR, EIEF, IZA. Email: ep454@cornell.edu

§Department of Economics, Cornell University. Email: qw98@cornell.edu. We thank Susan Athey, Patrick Bayer, Giacomo De Giorgi, Jessie Handbury, Tatsiramos Konstantinos, Mike Lovenheim, Michele Pellizzari, and Steve Ross for helpful comments.

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 sets differ substantially, as highlighted by the influential literature on information asymmetry (Akerlof, 1970; Spence, 1973; Rothschild and Stiglitz, 1976). In addition, information exchange and acquisition are costly and crucially depend on the interaction between the sending and receiving parties (agents, institutions, etc.).

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

In this paper, we focus on information's impact 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 the extent to which information flow is accompanied by worker flows. Second, we examine how information flow among friends affects job transitions and worker-vacancy match efficiency.

To 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 the course of twelve months. These detailed records enable us to construct measures of information exchange between geographic areas and among individuals, as well as variables on each user's employment status, history of work locations, home locations, and 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. Conditional on the amount of information exchanged, the 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 the information sources possessed by the working population is far more critical than the information sources possessed by the residential population. Surprisingly, in terms of the relationship between information diversity and economic development, our data exhibit remarkable similarity to the UK data analyzed by Eagle et al. (2010), 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 (job-related) information shared by friends on job switches. When an individual moves to a pre-existing friend's workplace, we define such a friend as 'a referral'.[2] 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 the job switch. In contrast, the information flow between workers and non-referral friends remains stable throughout the sample period, with no noticeable differences during the months that precedes job switches. The distinctions in mobile phone calling patterns are not driven by potential changes in the number of social contacts, which is steady throughout our sample period. These results provide suggestive evidence that individuals seek 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 employer relocate employees in different time periods, we might observe workers moving to the work location of pre-existing social contacts. Multi-plant firms are a rare phenomenon in the Chinese manufacturing industry; to the extent that this problem matters, we tackle it by adding the *interaction* of the origin and destination neighborhood fixed effects. Essentially, we compare individuals who share the same origin-destination neighborhood pair but have different social networks and examine their choices of workplace locations with and without friends.

The second confounding factor, a long-standing challenge in the literature that examines observational data, is the difficulty of distinguishing a referral effect from homophily and sorting.

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

[2] In our setting, referrals are social contacts in the workplace and include individuals who share job openings with friends as well as those who tell their employers about the attributes of their friends. We discuss these two different channels in Section 4.2.

If individuals share similar skills and preferences with their friends, then an individual might move to a location where a friend works not because of the referral information but because the vacant position requests certain skill sets. In addition, not all locations have desirable openings. Leveraging the richness and structure of our data, we conduct a battery of tests. First, we limit our analysis to individuals for whom there is at least one additional location within the same neighborhood that has vacancy listings in the same occupation as the occupation that the job switcher 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 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. friends who live at a location. Larger estimates for friends working in the location would be consistent with referrals: affiliation with the workplace enables friends working there to have an information advantage of job openings. Our results illustrate that friends currently work in the location are indeed much more important than friends who moved away prior to the job switch and friends who live there.

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

Job information passed on via referrals is valuable for workers. Specifically, referral jobs are associated with higher wages and non-wage benefits, shorter commutes, and a greater likelihood that workers 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. We also find suggestive evidence that referrals improve labor market efficiency by providing better matches between workers and vacancies, and, because women and migrants are more likely to find jobs through referrals, referrals mitigate labor market inequality.

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

Our work is related to the empirical literature on information economics. Recent studies have shown that increasing information transparency (for example, through better labels and postings) helps consumers' perceptions of product attributes (e.g., Smith and Johnson 1988), improves consumer choices (e.g., Hastings and Weinstein 2008), 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 that facilitate 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 of the diversity of socioeconomic outcomes and illustrate the importance role they play in shaping worker flows.

Another relevant strand of literature examines the role of social networks in job searches (Topa, 2011; Schumutte, 2016). To identify referred workers, this literature uses surveys or assumes interactions and exchange of job information between social ties, such as fellow workers, family ties, ethnic groups, residential neighbors, and Facebook friends.[3] The paper closest to ours is Bayer et al. (2008), who also study the importance of referral effects in an urban market. Using Census data on residential and employment locations, they document that individuals who reside in the same city block are more likely to work together than those who live in nearby blocks, and they interpret these findings as evidence of social interactions. We contribute to this literature by providing a superior measure of social networks and information exchange among individuals, and we introduce complementary data on vacancies and firm attributes to cover a diverse set of economic outcomes.

The paper proceeds as follows. Section 2 presents motivating evidence that information flow strongly correlates with the flow of workers. Section 3 discusses data and the institutional background. Section 4 presents the regression framework and reports results from the empirical

[3]In the existing literature researchers have proposed several proxies for social networks, such as former fellow workers (Cingano and Rosolia 2012; Giltz 2017; Saygin et al. 2018), family ties (Kramarz and Skans 2014), individuals who belong to the same immigrant community or ethnic group (Edin et al. 2003 ; Munshi and Rosenzweig 2013; Beaman 2012; Dustmann et al. 2016; Aslund et al. 2014), residential neighbors (Bayer et al. 2008; Hellerstein et al. 2011; Hellerstein et al. 2014; Schumutte 2015), and Facebook friends (Gee et al. 2017)

analysis. Section 5 concludes.

2 Motivating Evidence: Information Flow and Worker Flow

We are interested in understanding how information exchange affects urban labor markets. Faced with 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 dataset that contains the universe of phone records for all persons in a metropolitan city who subscribe to a major telecommunication operator in China. This dataset provides 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 to the worker flow constructed using the same sample.[4] To the best of our knowledge, this is the first 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 levels of spatial aggregation. At the highest level, the city is divided into twenty-three administrative districts. These districts are further broken into 1,406 neighborhoods that are delineated by major roads (or 'cells'), with each neighborhood populated by a varying number of locations.[5] In total there are close to eighteen thousand locations.

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; the width of each edge is scaled proportionately to the number of switches. Red non-directional edges denote the average number of weekly calls, and these have scaled edge-widths as well.[6] Note the remarkably strong correlation between 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 that have the thickest edges are, respectively, the commercial center of the city which has large retail chains, and an urban core with the second highest GDP among

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

[5] The average sizes for an administrative district, a neighborhood (cell) in the city proper (i.e., the urban center of the city), and a neighborhood (cell) in the suburb are 712 km^2, 0.45 km^2, and 25.03 km^2, respectively. A location is a building complex within a neighborhood (cell).

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

all areas in the city, respectively. The strong correlation remains when we include districts in suburbs that have less economic activity (fewer job switchers and lower call volumes).

Some correlation naturally arises from heterogeneous spatial and economic attributes, such as the two economic centers in the example described above. To address this, we run a regression analysis and control for origin and destination fixed effects. Regressing the worker flow between a 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. The R-squared is 0.24 when the number of calls is the only regressor, and it 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 our analysis to calls received or made by job switchers prior to their job change in Column 2. In practice, some calls might be initiated after individuals have decided to move and could reflect communications arising from newly established (work) relationships. In Column 3 and 4, we further exclude calls made within one month (Column 3) and three months (Column 4) of the job switch. When we exclude calls that might 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.

This strong correlation persists at finer geographical scales. 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 demanding exercise. Reassuringly, the positive correlation exists even at this fine scale, 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, indicate that job-related information flow plays an important role in worker flows.

Out-of-sample Prediction Existing studies have shown that mobile phone usage can predict economic activities. Following Kreindler and Miyauchi (2019), we use the uncovered relationship between information exchange and worker flow among neighborhood pairs during the first half of the sample to predict worker flow during the second half of the sample. We report a couple of measures of the accuracy of the prediction exercise. We first check the correlation between our prediction and the observed outcome. The higher this correlation, the better.

Then we regress the observed outcome on the predicted value and report the R-squared for the following 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 control for cell fixed effects), the out-of-sample prediction exercise does well. In all cases we have examined, the regression coefficient between our prediction and the observed outcome is close to one, varying between 0.97 to 1.03 depending on specifications. The R-squared varies from 0.30 to 0.32, which is high for cross-sectional studies with a large sample. The correlation between the predicted and actual worker flows is 0.55-0.56 across specifications. These encouraging results 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 economics literature have long emphasized the importance of diversity (Ottaviano and Peri, 2006; Ashraf and Galor, 2011; Alesina et al., 2016). In our setting, the content and value of information might vary over time and across individuals. 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 into 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] Social entropy measures the diversity of an individual's 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. It 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 the number of i's friends, is a scaling number that normalizes the Shannon entropy. Normalized entropy measures are guaranteed to vary from zero and one and are comparable across different measures, with higher values representing more diverse outcomes.

[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.

Spatial entropy measures the diversity of an individual's social ties in geographic locations:

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

where P_{il} is the probability of communication between individual 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 income entropy as:

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

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 defined as above. As in the other entropy measures, the normalization is measured 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 who reside or work in each location. A high value indicates that the working or residential population at a particular location communicates with diverse sources of information. To examine the importance of diversity, we regress the log of worker flow on the average entropy measures at the location level. Our controls include the total call volumes, which, as shown above, is an important predictor of worker flows; the number of individuals (subscribers of our data provider) observed in a location, which captures the scale effect (that is, more populated areas naturally have a higher job inflow); and 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 stacks all three measures together.[9] Both social and income entropy, which reflect the socioeconomic diversity of individuals' information sources, have a sizable and significant impact on job inflow *conditional on* the total number of calls, although the correlation between income entropy and worker flows is stronger. A one standard deviation increase in social and income entropy

[9]Here we limit our observations to locations that have 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 ten residents.

raises the worker inflow by 3% and 10%, respectively. Spatial entropy, on the other hand, is insignificant with a negative sign. This might be because our sample consists of individuals from the same city with limited spatial diversity.

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

It is worth noting that our results are remarkably similar to the findings in Eagle et al. (2010), who examines phone calls in UK in 2005 and relate 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 the UK, which reflects a denser social network in China), the average minimum number of direct or indirect edges that connect two individuals is very similar (10.4 in our context vs. 9.4 in the UK). Moreover, as in our setting, there is a strong correlation (varying from 0.58 to 0.73) between information diversity and the socioeconomic development of communities in the UK. These results reflect common features of the role of information diversity at play across different socioeconomic contexts, and it is unlikely that these features are limited to specific markets or time periods.

Having illustrated the high correlation between information exchange and job flows, we turn to the bulk of our empirical analysis that focuses on a specific channel of information at work: information on job openings shared among social contacts. The existing literature has documented that 30 to 60 percent of all jobs are typically found through informal contacts rather than formal search methods (Topa, 2001; Burks et al., 2015); this is a universal pattern that holds across countries, over time, and regardless of occupation or industry. Our calling data provide superior coverage of information flow through social networks. We also bring in a large number of complementary data sets. 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 assembled auxillary 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 of 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 records activities of all of Company A's subscribers in the city we study.[11]

Unlike social media, the cellphone penetration rate is very high in China. According to the China Family Panel Studies, a nationally representative longitudinal survey of individuals' social and economic status since 2010, 85% of correspondents sixteen years and older report possessing a cellphone.

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

An important advantage of our data is that it includes geocoded locations at the time of use. Whenever a user activates his mobile device, the serving tower station records a geographical position in longitude and latitude that is accurate up to a 100-200 meter radius, or roughly the size of a large building complex. For each individual and week, we observe the location that has the most frequent phone usage between 9am and 6pm during the weekdays (which we call a 'work location') as well as the location that has the most frequent usage between 10pm and 7am for the same week (which we call a 'residential location').[12] These geocoded locations trace out individuals' spatial trajectories over time, allow us to construct diverse types of social ties (including friends, neighbors, past and present coworkers, friends' coworkers), and are unique among social science studies, which typically lack fine-grained geographical information about human interactions.

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

[11] For individuals with multiple phones, we observe usage on the most commonly used phone. If they subscribe to services from multiple carriers (which is uncommon), we only observe activity within company A.

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

Constructing individuals' workplaces histories using recorded geocodes is the most crucial step of our analysis. Since we do not directly observe employment status or place of work, we take a conservative approach that mitigates measurement errors in our work-related variables. We focus on individuals with valid work locations for at least forty-five weeks – a period long enough to precisely identify job switches. This gives us 560k individuals.[13] After further restricting our sample to individuals who have at most two working locations throughout the sample period (which excludes sales persons and people with out-of-town business travels and family visits) and for whom we have complete demographic information, our final sample reduces to 456k users. We carry out the core empirical analysis using this sample and conduct robustness checks in Section 4.5 using less stringent sample selection criteria.

We identify individual i as a *job switcher* if the following criteria are satisfied. First, as shown in Figure 3, a job switcher is someone who has two work locations and is observed in both locations for at least four weeks each. Locations that are visited during the working hours on a daily basis for weeks in a row are likely to be a work location rather than shopping centers or recreational facilities. Second, the distance between these two locations must be at least 1 km. To reduce the possibility of erroneously identifying someone as a switcher, we choose the cutoff of 1 km because 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 in Western countries (e.g., 15-18% in the European Union as documented in Recchi 2009), partly because of the Hukou system, which imposes significant restrictions on individuals' moving across provinces or from rural to urban areas (Ngai et al., 2017; Whalley and Zhang, 2007).

The city we study is divided into 1406 neighborhoods, or cells, that are delineated by major roads. A cell is similar to but is smaller in size than a census block in the U.S. There are 917 cells in the city proper (i.e., the urban center of the city) and 589 cells in surrounding counties (see Figure 2 for a section of the city map)[15]. A lower level of geographical unit is a *location*, or the geographic position returned by a tower station, which represents a building complex or an establishment within a cell. The median and average number of distinct locations in a cell is seven and thirteen, respectively. Our switchers found jobs in a total of 5,800 work locations that are spread in 1,100 cells, about two-thirds of which are in the city proper; the reminder

[13]Several factors contribute to sample attrition. China's cellphone market is dynamic, with a high fraction of subscribers switching carriers during each period. In addition, the 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 (especially toward the end of a billing period for subscribers on prepaid plans).

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

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

are in surrounding counties.

Vacancy Data High quality data on vacancies and matches with workers are extremely hard to find. To gauge the dynamics of local labor market conditions, we collect listings from the top two largest online job posting websites, zhilian.com and 58.com, from August 2016 to Feburary 2018.[16] These websites hold on average 10,000 job postings per month. We obtained a total of 121,055 postings and merge them to our call data based on locations.

Each posting reports the posting date, job title and description, full time or part time, qualifications (minimum education and years of experience), monthly salary (in a range), firm address, firm size (number of total employees), and firm industry. On the basis of the job title and description, we group these postings into eight occupations using the 2010 U.S. occupation code. Popular occupations include Professionals (26.70%), Service (26.61%), Sales and Office administration (19.24%), and 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., an annual salary of RMB 25k-40k). Using the mid-point of the reported salary range delivers a rather flat wage profile across industries: jobs in the construction sectors are entitled a salary that is similar to jobs in professional services. Missing salary information in postings is also common. Finally, a sizable fraction of workers' compensation consists of non-wage benefits, including bonuses and commissions, paid vacations, health and unemployment insurance, etc. (Cai et al., 2011).

We utilize two firm-level administrative datasets to obtain wages and benefits, local industry composition, and firm attributes. The first is the annual National Enterprise Income Tax Records from 2010 to 2015, which are collected by the State Administration of Taxation, which contain firm ID, industry, ownership, balance sheet information (revenue, payroll, employee size, etc), and 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. Our final data set contains firm location, industry, ownership type (whether or not state owned), employee

[16] 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`).

size, revenue, wage payroll, and capital for a total of between five to ten thousand firms.[17]

In our sample most 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, which is 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, we find 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 RMB 32,005 vs. RMB 20,609).

Housing Price Our main data source does not contain individuals' socioeconomic measures such as wealth or income. To overcome this data limitation, we scrape housing data from Anjuke.com, a major online real estate brokerage intermediary and rental service provider in China that collects housing information for both residential and commercial properties. For each residential complex, Anjuke.com reports its name, property type and attributes, the monthly average housing price per square meter, year built, total number of units, average size, and street address. About 64% of the cells in the 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 occupations among job postings, industry composition, number of employees and vacancies, average wage, and housing price. For each individual in our final sample, we observe his work and residential location, friends, neighbors, friends' workplaces and home locations, as well attributes for each location.

Chinese Labor Market China's labor market has several noticeable features. Relative to other developing countries, China has a high female labor participation rate. In response to the employment pressure generated by its large population, China has instituted a mandated (early) retirement age, 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 on the basis of their place of birth, and its goal is to anchor peasants to the countryside. According to Zhang and Wu (2018), China's urban labor market has a two-tier system: urban cities and rural areas. The large divide that separates these two tiers in terms of job opportunities, social benefits, and amenities (education, health care, etc.) has created a high fraction of migrant workers in urban cities who take jobs with low wages and long working hours and often are denied social benefits.

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

State owned enterprises (SOEs) account for a small fraction of the total number of firms, but they constitute approximately 30 to 40 percent of China's GDP and 20 percent of total employment (State Assets Supervision and Administration Commission 2017). Many SOEs appear in the Fortune Global 500 list and are among the largest conglomerates in the world. 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 wages, which are sometimes higher wages than those in non-state sectors.

Similar to the 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; 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, where a higher fraction of young correspondents cite referrals as their main channel of landing a job.

Summary Statistics: Demographics and Social Ties Table 5a presents descriptive statistics of the 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; the rest migrated from other provinces. Thirty-nine percent of users are born in the city proper.

The bulk of our analysis focuses on job switchers and their social networks. 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.[18] As Table 5b illustrates, job switchers bear demographics similar to those of non-switchers, except for age. Job switchers are more likely to be in their thirties and on average are two years younger than non-switchers. They are less likely to be migrants and have a smaller fraction of friends who use Company A's mobile service, although the magnitude of these differences is 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 concerned about potential sample selection bias if Company A's subscriber network over-represents certain demographic groups. This is unlikely to be a

[18] They are also called 'one-way' contacts. An alternative definition requires a contact to make a phone call to *and* receive a phone call from individual i at lease once during the sample period. These two definitions lead to very similar results. Section 4.5 conducts robustness checks on the definition of social contacts.

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. To test the robustness of our results with respect to potential sample selection bias, Section 4.2 separates individuals with the friend coverage above the median from those with coverage below the median and documents similar findings.

To help interpret the magnitude of the coefficient estimates in Section 4 below, Appendix Table A2 tabulates the summary statistics for key variables referenced in various regression samples.

4 Empirical Analysis: Referral-Based Worker Flow

Throughout the section (except when noted otherwise), we limit an individual's network to the one formed prior to his job switch [19]. This avoids endogenous links formed after the job switch.

Among the 38,102 job switchers observed in our sample, 4,703 (12%) of them have missing friend locations (Panel A of Table 6). Among the switchers with non-missing locations 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 have forty-five weeks of non-missing work locations.[20] As discussed in Section 3, forty-five weeks ensures the accuracy of identified job switches, but it 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, while Section 4.5 repeats these analyses using friends for whom there is at least four weeks of work information. The results are robust to this alternative friend definition.

One might be concerned that the referral definition in our sample suffers from confounding factors, in particular, sorting or homophily. We proceed as follows. We first conduct an event study on the time series variation of information exchange between job seekers and referral vs. non-referral friends in Section 4.1. Then we perform a battery of regression analyses to illustrate that our estimated referral effect is not driven by confounding factors in Section 4.2. Finally, we evaluate the benefits of referrals to workers in Section 4.3 and benefits to firms in Section 4.4.

[19] Our analysis has also been performed when excluding social contacts formed up to three months prior to the job switch since those might be formed for reasons related to the move. Results remain qualitatively unchanged.

[20] In addition, we do not observe other types of networks (relatives, classmates, social contacts via WeChat, etc.); hence our results are necessarily a lower bound.

4.1 Event Study

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

We first show that individuals' weekly number of contacts prior to their job change is stable. Using all of job switchers' friends (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 varies between twenty-three and twenty-five for most weeks, with a modest decrease just prior to the switch. This suggests 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 contacts communicated with after the job switch is moderately higher, which is intuitive and reflects new relationships formed at the new work place.

To examine the dynamics in information flow, we regress the phone call frequency between individual i and his friends on the event window of eleven months before and ten months after the job switch, with a rich set of fixed effects:[21]

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

where Freq_{ijt} is the number of calls between caller i and his friend j in month t, Referral_{ij} takes value one if switcher i moves to friend j's workplace during the sample period and zero otherwise, Non-Referral_{ij} takes value one when Referral_{ij} is zero, λ_i is an individual fixed effect, and τ_t is a month fixed effect. The key coefficients $\{\gamma_s, b_s\}$ vary by event month s relative to when the job switch occurs ($s = 0$ for the month of switch). 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 b_s because switcher-non-referral pairs are more common: there are 4.9 million switcher-non-referral-month observations relative to 253k switcher-referral-month observations.

The communication patterns between referral and non-referral pairs are distinct, even after controlling for a rich set of fixed effects. First, switchers have more frequent calls with referral friends than non-referral ones. Second, the intensity of information flow between switchers 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 prior to the job switch. Importantly, non-referral friends also include social contacts who live in the area but not work there. If all the communication

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

were about local amenities one would expect no substantial difference between the two curves. 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 dimensions of social interaction, as referrals and referees are friends before the job switch and become friends and colleagues afterwards.

One might be concerned that individuals might share with friends news about their job offer, which would also lead to intensified communication before they move to the new work place. 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. Finally, some phone calls between the referral pairs could be inquires about workplace amenities instead of job openings. We regard all such calls as communication via referrals that facilitates a job switch.

4.2 Referrals and Work Location Choices

We turn to a regression framework to quantify the degree to which the referral effect shapes 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 a nearby location, we use the following model:

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

where 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, while λ_c denotes cell fixed effects that control for unobserved location attributes (number of job vacancies, industrial composition, number of locations, etc.). Demographic controls $X_{ki},\ k = 1, ..., K$ consist of 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 it lies 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 purposefully. Job location choices are influenced by many factors, including industry composition and local labor demand, commuting distance, local amenities, intra-household bargaining for married couples, many of which cannot be directly controlled in our framework. Limiting an individual's 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 who move 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 wherein the key coefficient β is estimated from the within origin-destination variation. We essentially compare individuals who have the same origin-destination cell pair but different friend networks and examine their choice of locations in the same neighborhood.

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

Leveraging the richness and structure of our data, we propose the following battery of tests. First, we limit our analysis to workers for whom there is at least one other location within the same cell that has vacancy listings in the same occupation as the one that he takes.[22]. This mitigates the concern that individuals sort into friends' locations that provide the only employment opportunity in the area.

Second, we distinguish between friends who are currently working in location l and friends who used to work there but moved away prior to the job switch. Given that sorting by unobserved preferences or skills should happen regardless of a friend's *current* location, we would expect to find similar β estimates for both types of friends if our finding is driven by sorting. Third, we distinguish between friends who work vs. friends who live at location l. Larger estimates for friends who work in location l would be consistent with the referral effect: affiliation with the workplace enables friends who work there to have an information advantage of jobs

[22]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.

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, a dummy for each age group, migrant status based on the birth county, and the total number of friends prior to the job switch.[23] Columns 3 and 4 repeat the first two columns with cell fixed effects for the new workplace. Columns 5 and 6 further include 21k fixed effects for the pair of old and new work cells.

The mean propensity to choose a location within a cell among the switchers is 0.09. The coefficient for the referral effect, which ranges from 0.34 to 0.36, is economically large, precisely estimated, and stable across all columns in Table 7. 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 has little impact on the key parameter estimate.

In Table 8, we conduct a goodness-of-fit exercise similar to that performed in an independent study by Buchel et al. (2019), and we report the percentage of correct predictions (the second to the last row). A correct prediction is one in which the observed location choice has the highest fitted linear probability. Column 1 only includes pair fixed effects. Column 2 adds Friend_{il}. Column 3 further controls for the number of calls between individual i and location l prior to the job change, echoing results documented in Section 2. Adding the friend dummy in Column 2 boosts the R-squared by 2.5 times – from 0.07 to 0.14 for a sample of nearly one million observations – which is remarkable. Correspondingly, the fraction of correct predictions is 8.9% in Column 1 and jumps to 23.9% before further increasing to 30% in Columns 2 and 3, respectively.

One might be concerned about sample selection bias given that information about work location is missing for friends outside Company A's subscriber network. Table 9, which splits the sample based on whether the friend coverage is above or below the median (the cutoff is 48%), replicates Columns 2, 4, and 6 in Table 7. The difference in the friend coefficient is modest (smaller than 0.02) and insignificant for our preferred specification.

To evaluate whether our finding is driven by sorting, we conduct in Table 10 the three tests described above. All columns include the old and new work cell-pair fixed effects and demographic controls. Columns 1 and 2 are limited to the subset of switchers who have at least one alternative work location within the same cell that has openings in the same occupation as the one the switchers take. This modestly impacts on the estimate: the coefficient of Friend_{ij} changes from 0.35 to 0.34. Columns 3 to 6 use the same sample as that in Columns 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. friends living there. In both

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

cases, friends currently working in the new location have a much larger impact on the choice probability: they are four times more influential than friends who recently moved away and 150% more effective than friends who live 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 operates 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 different mechanisms, we test their common implication that referrals mitigate information frictions in the hiring process. We thus examine whether referrals are more important when information asymmetry is more severe.

Individuals who live far away from the new work location, have limited work experience, or change industrial sectors are likely to be disadvantaged when it comes to obtaining information about new openings. Similarly, employers are less likely to be knowledgeable about these workers. In Table 11, we interact Friend_{il} with the distance between the old and new work place, the distance between home and the new work location, a dummy for young workers (between 25 and 34), moving from rural to urban locations, and changing sectors.[24] 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 0.68 and 0.53, respectively, which is a significant boost above the base estimate of 0.35. In Column 7 of Table 11, we interact Friend_{il} with the demeaned number of calls between individual i and friends in location l prior to the job-switch. The referral effect increases with calling intensity: one hundred calls are associated with a two percentage point increase in the probability of moving to a friend's place, which is consistent with findings in Gee et al. (2017), who examines the effect of strong social ties on job search using Facebook friends.

The evidence in Table 11 also allow us to rule out two alternative explanations for our findings. One is that our results are simply driven by preferences: individuals enjoy the company of friends and hence prefer to work at their place. However, the stronger referral effect when information asymmetry is more severe, documented in Table 11, is inconsistent with this possibility. Similarly, could our estimates be mostly driven by nepotism, that is, friends and family are hired instead of the best available candidates (Hoffman, 2017)? It is probably not so in our

[24]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.

case: such estimates would not predict that the effect is stronger when information asymmetry is more severe. Moreover, as shown below, referrals are more common among people in the same age range, whereas nepotism often involve individuals from different age groups (children of relatives) (Wang, 2013; Foley, 2014).

Comparison with the Literature How do our results compare to the existing literature that examines job referral effects? There are two common approaches of inferring referrals in observational studies. The first, pioneered by Bayer et al. (2008), defines referrals as residential neighbors. Using data from the Boston metropolitan area, they consider as friends individuals who live in the same Census block. The second approach assumes that social interactions are stronger within the ethnic group and it defines friends as co-workers who are members of 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 and works in location l. Ethnicity, which is inapplicable in China's context, is replaced with birth county as the literature documents strong social ties among individuals from the same birth region (Zhao, 2003).[25] '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 who are not neighbors, while Column 4 compares coworkers who share the same birth county with friends who work in the same location but have different birth counties.

The results shown 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 the same birth county are 0.21 and 0.10, respectively, when they are the only measure of an individual's social network. Given the average moving probability of 0.09, having a social tie of either type more than doubles the probability of switching to location l. On the other hand, friends dominate both types of social ties by a large margin. The difference in magnitude is both statistically significant and economically sizable, and in the case of 'same birth county', the effect of friends is three and half times as large. Our results confirm results in the literature but suggest that they constitute a lower bound.

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 wherein the probability that individual i moves to friend j's workplace is a function

[25] In China, counties are a lower level of geographical unit and smaller than cities. Seventeen counties comprise the city we study.

of worker 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 M_{ij} is one if i moves to friend j's workplace and 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, the 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 $\{i, j\}$, wherein i moves to j's work location, with dyad $\{i, m\}$, where i does not move to m's work location.

We limit the regression sample to the subset of switchers (8,518 individuals) who find a job at some friend's workplace. Column 1 of Table 13 includes all eligible dyads that have non-missing demographic information, for a total of 93k observations. Column 2 only includes switchers for whom there is at least another location within the same cell that has vacancy listings in the same occupation as the one that the mover takes; here there are 88k observations. Females and migrant workers are more likely to receive referrals. There are strong assortative patterns in referral provision. Females on average are less likely to provide referrals but they are more likely to provide referrals to other women. Similarly, workers are more likely to refer other workers who are from the same hometown county. This is consistent with recent findings that in China community networks based on birth county facilitate entry and the growth of private enterprises (Dai et al., 2018). Finally, older workers are more likely to provide referrals, whereas individuals of similar age are more likely to refer jobs to each other, although 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

In this section, we examine whether referrals improve referees' labor market outcomes, conditional on finding a job. Our framework for analyzing 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 who live in residential cell r and switch to work location l in cell c. We control for the same set of demographic variables considered in

model (1). Because we do not observe individuals' socioeconomic background, such as education and wealth, we include in all regressions the residential cell fixed effect (α_r) as a proxy for one's socioeconomic status (luxurious complexes vs. low-income neighborhoods).

We construct five different measures of job quality. Our first measure is the expected wage at the new job, measured by the average annual payroll (in thousand RMB) among firms in the same location weighted by their number of employees.[26]

Wage dispersion is primarily driven by across-firm rather than within-firm differences (Card et al., 2018) and an individual's housing value is correlated with his labor income. Thus, we use *coworkers'* housing price as a second measure to proxy for monetary compensation. Specifically, we construct the difference between the 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 wages and other pecuniary benefits.

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

Results Because 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 RMB 620, or about 2% of the average wage reported in our sample.[27] Turning to differences in coworkers' home values in the new vs. the 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 RMB 13,000 ($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 2% increase in the likelihood of working full-time.[28] Thirty-one 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,

[26]We assign each firm in the tax data 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.

[27]The annual wage is measured in thousands RMB and the mean is 31.

[28]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 the hours worked is six. This is likely to be an under-estimate of the actual hours worked. Part-time (full time) is defined as thirty hours or less (more than thirty hours). On average, 57% of the switchers work full-time before the job change, reflecting the conservativeness of our measure.

having a friend at a SOE firm raises the probability of moving there by one percentage point, which is a 9% increase from the mean (0.11).[29] To the extent that higher wages are an indication of worker productivity and shorter commutes and full-time positions reflect better job amenities, our results provide evidence that referrals lead to better matches between workers and vacancies and improve workers' labor market outcomes through mitigating information frictions in the hiring process.

4.4 Referral Benefits To Firms

With a few exceptions, most empirical studies of job referral effects abstract away from analyzing firm outcomes because comprehensive data on the performance of both employees and employers are hard to obtain.[30] 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 merge 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 that have more than one hundred employees, which represent about 20% of our sample. While this choice significantly reduces the sample size, it mitigates measurement errors because there could be multiple firms in the same location and it is difficult to match workers to firms. The average employment for these large firms is 150; thus, it is likely that these firms occupy an entire location, and consequently, reduce the likelihood of erroneously linking workers to unrelated firms. In the rest of this section, we use "location" and "firm" interchangeably. Appendix Table A3 reports results from replicating the analysis using all firms. The referral effects are similar both statistically and economically, which is reassuring.

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

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

where i denotes a firm. We examine three measures of firm performance (Y_i): (1) net inflow of workers; (2) match rate, measured by the number of hires over vacancies; and (3) firm growth rate, measured by the number of hires over total number of employees.[31] We use $log(Y+1)$ for the net inflow to include observations that have 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

[29] A workplace is classified as 'SOE' if the majority of workers at that location are employed by SOE firms.

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

[31] Observations with negative net inflow are dropped.

our analysis to locations with at least one hiring, 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 who switches to firm i has a friend working there, while λ_c denotes cell fixed effects – the same as in model (3). Variables $\{Z_{ki},\ k = 1, ..., K\}$ denote firm attributes and employee characteristics. Firm attributes include firm age, the average number of employees (firm size), dummies for eighteen different industries, large firms, and SOEs, and average real capital from 2010 to 2015. To capture pre-existing trends, we also control for the average employment growth rate from 2010 to 2015. In addition, we include a firm's referral network size, which is defined as the number of unique social contacts owned by employees who work in firm i prior to the arrival of the new hires.[32] Worker attributes include the shares of female workers and migrants, the average age of employees, and average housing price of the pre-existing employees.[33]

Results The parameter estimate γ captures the effect of using referrals on firms' performance. To the extent that firms that grow quickly are more likely to hire through employee referrals, our estimate could be biased upward. To tackle this problem, we estimate model (4) by increasing the set of variables that help to control for firm growth and employee quality. Nonetheless, because we lack suitable instruments and only have a cross-section of data, our results in Table 15 are largely descriptive.

The Referral_i coefficient estimates are remarkably similar across different sets of controls for firm and worker attributes, indicating that our results are unlikely to be inflated by unobserved firm or worker characteristics. Firms that hire 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 73%, enhances the job matching rate by 86% (the average matching rate for large firms is 0.76), and raises the firm growth rate by 45% (the median growth rate is 4% for large firms). Results in Appendix Table A3 document similar patterns. Although our analysis in this section is descriptive, the fact that the estimates are robust to a detailed set of firm and worker controls raises our confidence that these estimates are not simply picking up unobserved characteristics related to firm and employee quality.

4.5 Alternative definition of friends

We conclude our analysis with a few additional robustness checks. Our core analysis is thus far limited to friends who have at least forty-five weeks of non-missing work locations. This mitigates measurement errors in job locations (which alleviates the attenuation bias) but omits a

[32] We exclude contacts established after the job switch.

[33] The number of Company A's users at each location is included in all regressions.

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$_{il}$', could be biased toward zero. In this section, we examine the robustness of our results to alternative sample selection criteria.

Appendix Table A4 replicates Table 7 yet includes all friends who have at least four weeks of non-missing work locations. This enlarges the number of individual-friend pairs from 401,437 to 979,595. The estimated referral effect remains unchanged: having a friend in a location increases the probability of moving there by 35 percentage points.

Using this alternative definition of friend, referral jobs are associated with a 1.3% increase in wage premium, a 0.6% increase in job-related benefits (as proxied by coworkers' housing prices), and a 12% increase in the likelihood of working full-time (Appendix Table A5). These effects are similar to those found in our base specification. The effect on the likelihood of a shorter commute and transitioning to a SOE firm is nearly identical to that found in the base specification. Turning to the referral benefit for firms, the alternative definition of friend produces slightly more pronounced results than those reported in the base specification (Appendix Table A6). We have replicated our analysis with various other friend selection criteria (e.g., including all friends who have at least three months or six months of work locations) and obtained very similar results.

Finally, Appendix Table A7 repeats Table 7 but defines individual i's friends as those who both place at least one call to and receive at least one call from individual i (social contacts with two-way communications). In addition, friends have at least four weeks of non-missing work locations. The referral estimate is again very similar to that of our base specification.

5 Conclusion

This paper uses novel geocoded mobile phone records to study how information provided by social contacts mitigates information asymmetry and improves the labor market performance.

Our study provides three broad lessons for future research. First, panel data with fine spatial and temporal variation hold great potential for overcoming the challenges of causal inference with observational data related to networks and markets. For example, in our context, the ability to identify different types of social contacts in small geographical areas at overlapping periods helps us tackle one of the most difficult challenges in empirical studies of referrals: sorting. Second, big data from non-conventional sources complement traditional data sources with socioeconomic outcome measures. In our analysis, merging in tax records and the business registry is crucial to shedding light on how referrals benefit firms, a topic that is understudied in the existing literature. Third, information exchange, and in particular, social and socioeconomic diversity in communication help to shape job movements. In the future, studies on the exact

mechanism that governs how information exchange through referrals increases labor market efficiency would be extremely valuable.

References

Abramitzky, R. and 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, NBER.

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

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, U. Schönberg, and H. Brücker (2016). Referral-based Job Search Networks. *The Review of Economic Studies 83*(2), 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.

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

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.

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

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.

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.

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

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.

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

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

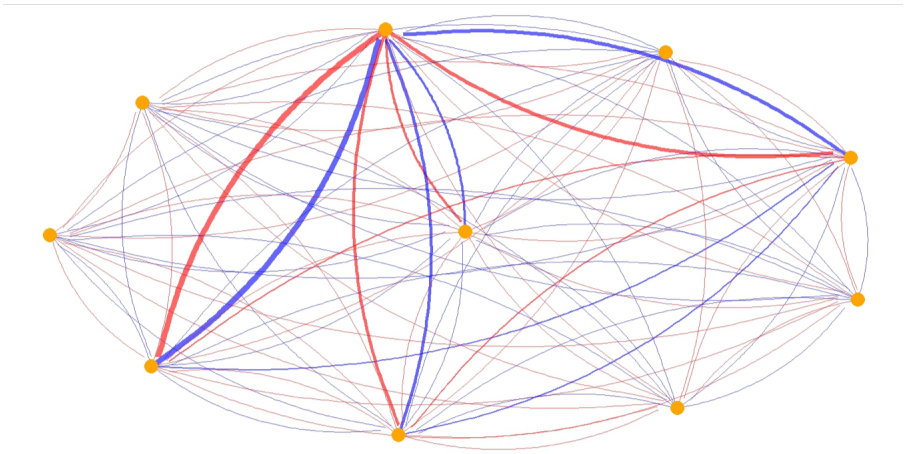
Zhang, J. and J. Wu (2018). The chinese labor market, 2000–2016. *IZA World of Labor* (437).

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.

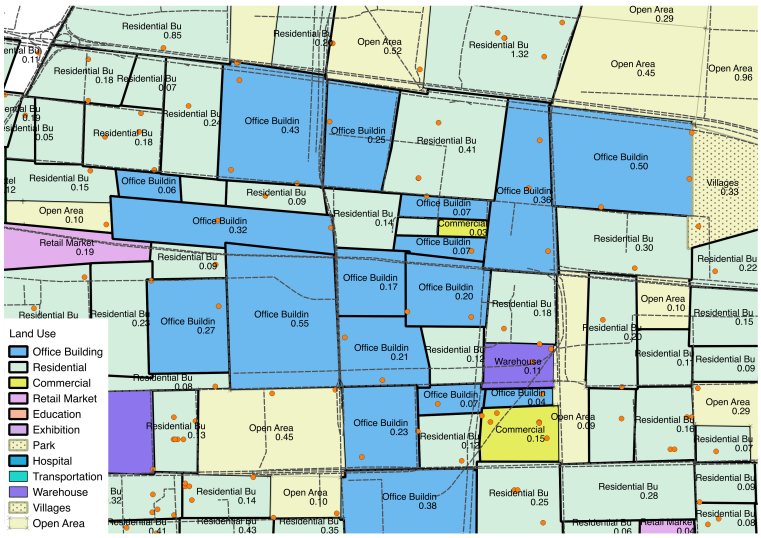
Figures and Tables

Figure 1: Information Flow and Worker Flow Among Administrative Districts



Notes: Each node is an administrative district in the city. We plot randomly selected ten nodes out of a total of 23. Blue (non-directional) edges 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 edges 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: the city is divided into 1,406 cells that are delineated by major roads (polygons in the map that are separated by dark lines) and close to 17,881 locations (orange dots). The number in each polygon denotes the area size in km^2.

Figure 3: Job Switch Timeline

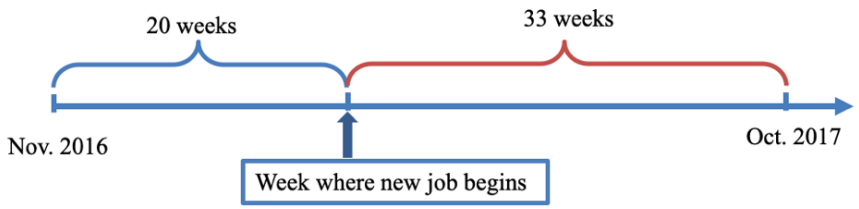
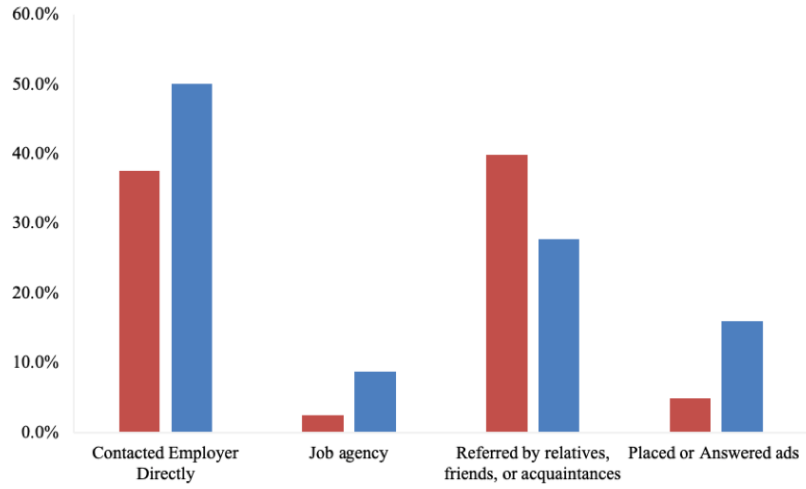
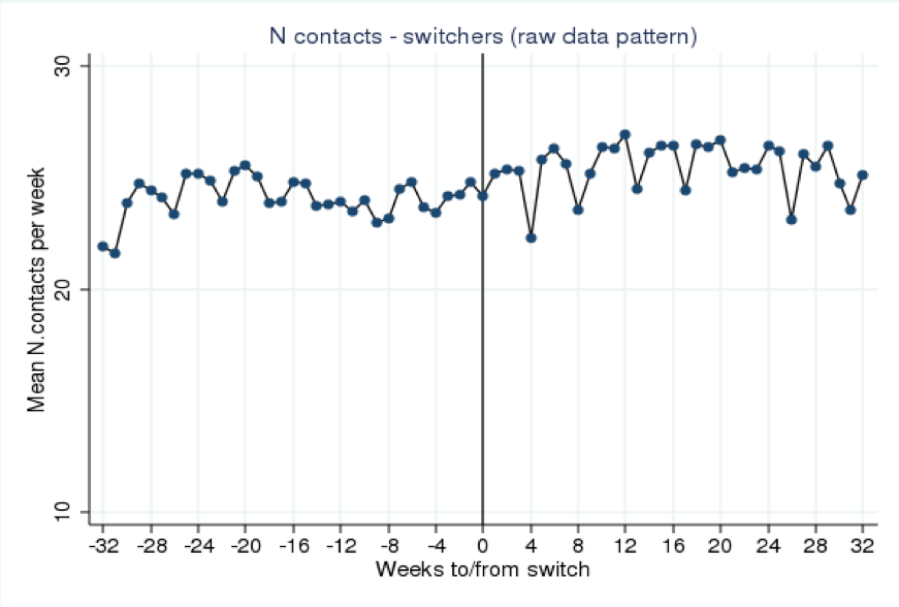


Figure 4: Job Search Methods in China vs. in the U.S. (2014)



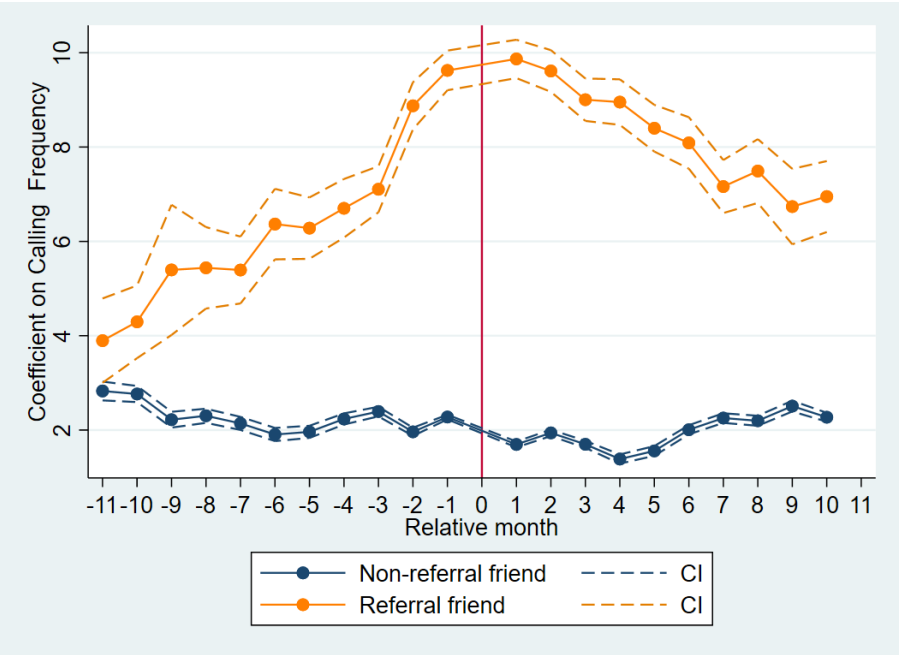
Notes: the horizontal axis reports different job search methods. The vertical axis displays the fraction of each method used among job seekers. Red (blue) bars represent China (U.S.). Source: China Family Panel Studies (2014) and U.S. Current Population Survey (2014).

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. The vertical line indicates the week of job switch. There are 37,099,345 switcher-friend-week observations.

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



Notes: orange line represents calls between switchers and the referrals (Obs = 252,852). Blue line represents calls between switchers and non-referrals (Obs = 4,915,656). Switcher fixed effects and month fixed effects are included in the regression.

Table 1: Information Flow and Worker Flows

(a) At the Administrative District Level

	All calls	Calls from/to job switchers ***before*** switch		
Dependent variable: Worker flow (i, j)		No exclusion	Excluding calls within 1 month of job switch	Excluding calls within 3 months of job switch
	(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

	All calls	Calls from/to job switchers ***before*** switch		
Dependent variable: Worker flow (i, j)		No exclusion	Excluding calls within 1 month of job switch	Excluding calls within 3 months of job switch
	(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). In Panel (b), one unit of observation is a pair of locations (i, j). There are 23 administrative districts and 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 columns 2 to 4, it is the total number of calls between switchers and their pre-existing contacts. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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

	Dependent variable: actual worker flow between (i, j)					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted worker flow (linear regression)	1.02***	1.02***				
	(0.18)	(0.002)				
Predicted worker flow (linear spline)			0.97***	0.97***		
			(0.16)	(0.001)		
Predicted worker flow (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
Num. 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. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The training data consist of switches in the first six months. The prediction is based on switches in the second six months. The table reports OLS regressions of the actual worker flow between cell i and j on the predicted worker flow. All three prediction models include neighborhood i and neighborhood j fixed effects. The linear prediction model uses the number of calls by switchers prior to the job change as a key predictor. The linear spline model uses five knots and the cubic spline model uses six knots, both are default options from STATA.

Table 3: Information Diversity and Worker Flows

Dependent variable: log 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
Num. of Cell FE	1,183	1,183	1,183	1,183

Notes: One unit of observation is a location with at least five workers and five residents. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Log inflow" is the log of the number of people moving to a given location. Social entropy, spatial entropy, and income entropy are normalized Shannon entropies as defined in the 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

Dependent variable: log 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
Num. of Cell FE	1,183	1,183	1,183

Notes: One unit of observation is a location with at least five workers and five residents. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Log inflow" is the log of the number of people moving to a given location. Social entropy, spatial entropy, and income entropy are normalized Shannon entropies as defined in the 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	435,098
Age 25-34	0.29	0.46	455,572
Age 35-44	0.26	0.44	455,572
Age 45-59	0.27	0.45	455,572
Age above 60	0.11	0.32	455,572
Age (midpoint)	40.18	11.97	435,194
Born in local province	0.75	0.43	455,572
Born in local city proper	0.39	0.49	455,572
Frac of social contacts in Company A	0.50	0.19	455,572
Job switch	0.08	0.28	455,572

(b) Switchers vs. Non-switchers

	Non-switchers			Switchers				
	Mean	SD	N	Mean	SD	N	Diff.	t-stat
Female	0.36	0.48	398,742	0.36	0.48	36,356	-0.00	-0.45
Age (midpoint)	40.36	12.00	398,817	38.23	11.49	36,377	2.13***	32.49
Born in local province	0.75	0.43	417,470	0.74	0.44	38,102	0.01***	3.62
Born in local city proper	0.39	0.49	417,470	0.38	0.49	38,102	0.00	0.70
Frac of social contacts in A	0.50	0.19	417,470	0.51	0.19	38,102	-0.00	-0.53

Notes: The sample is restricted to individuals 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 the fraction of individuals' contacts who are company A's customers. 'Job switch' is a dummy for job switchers, who are identified based on the criteria described in the text. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Percentage of Job Switchers Switching to A Friend's Workplace

Panel A: including friends with at least 45 weeks of location information			
	Percent	N. Individuals.	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: including friends with at least 4 weeks of location information			
	Percent	N. Individuals.	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: Job switchers are identified based on the criteria described in the text. Panel A includes all friends with a work location for at least 45 weeks. Panel B includes all friends with non-missing work locations 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" is the number of switcher-friend pairs where friends only include social contacts existed prior to the job switch.

Table 7: Referral Effects on Job Switchers

Dependent variable Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.36***	0.36***	0.34***	0.34***	0.35***	0.35***
	(0.003)	(0.004)	(0.02)	(0.02)	(0.01)	(0.01)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,151,676	1,120,797	1,151,676	1,120,797	1,151,676	1,120,797
R-squared	0.08	0.08	0.14	0.13	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
Num. of Cell FE	NA	NA	1,111	1,107	21,250	20,811

Notes: One unit of observation is a switcher-location pair. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Friend" is a dummy variable that 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 pre-existing social contacts (irrespective of carriers).

Table 8: Referral Effects on Job Switchers: % of Correct Predictions

Dependent variable Probability i switches to location j	(1)	(2)	(3)
Friend		0.35***	0.30***
		(0.01)	(0.01)
Num. calls to L (x1000)			0.39***
			(0.06)
Controls	Yes	Yes	Yes
Observations	1,120,797	1,120,797	1,120,797
R-squared	0.07	0.14	0.15
Old x New Cell FE	Yes	Yes	Yes
Num. of Cell FE	20,811	20,811	20,811
Correct predictions at location level	8.9%	23.9%	30.0%
Percent increase w.r.t previous column		170%	25.5%

Notes: This table replicates Column 6 of Table 7 using model (1), except that Column 1 excludes the "Friend" dummy and Column 3 adds "Num. calls to j", the number of calls (in thousand) between switcher i and location j prior to the job switch. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. A correct prediction is one where the chosen location has the highest fitted linear probability.

Table 9: The Referral Effect: by Friend Coverage

Dependent variable Probability i switches to location j	(1) Above	(2) Below	(3) Above	(4) Below	(5) Above	(6) Below
Friend	0.35***	0.37***	0.33***	0.35***	0.34***	0.36***
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Observations	612,230	508,567	612,230	508,567	612,230	508,567
R-squared	0.09	0.06	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
Num. of Cell FE	NA	NA	1,050	1,033	11,889	10,787

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. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: The Referral Effect: Falsification Tests

Dependent variable	Individuals with similar job opportunities nearby					
Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.34***	0.34***	0.34***	0.34***		
	(0.015)	(0.015)	(0.015)	(0.015)		
Friend moved before the switch			0.08***	0.08***		
			(0.021)	(0.022)		
Friend working there					0.30***	0.30***
					(0.013)	(0.013)
Friend living, not working there					0.19***	0.19***
					(0.013)	(0.013)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,134,849	1,104,171	1,134,849	1,104,171	1,134,849	1,104,171
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
Num. of Cell FE	20,062	19,644	20,062	19,644	20,062	19,644

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. Columns 3 and 4 compare friends currently working in the new workplace with friends who moved away prior to the job switch. Columns 5 and 6 contrast friends working there with friends living there. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Referral Effects and Information Asymmetry

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Probability i switches to location j							
Friend	0.35***	0.33***	0.33***	0.33***	0.34***	0.26***	0.34***
	(0.014)	(0.015)	(0.016)	(0.015)	(0.014)	(0.015)	(0.014)
Friend×Distance(job1, job2)		0.002***					
		(0.0003)					
Friend×Distance(home, job2)			0.004***				
			(0.001)				
Friend×Young (Age 25-34)				0.04***			
				(0.009)			
Friend×Rural to urban					0.34***		
					(0.039)		
Friend×Change sector						0.27***	
						(0.022)	
Friend×Call intensity							0.0002***
							(3.22e-05)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,120,797	1,120,797	1,041,950	1,120,797	1,120,797	240,435	1,120,797
R-squared	0.14	0.14	0.15	0.14	0.14	0.15	0.14
Old x New Work Cell FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of Cell FE	20,811	20,811	19,595	20,811	20,811	5,684	20,811

Notes: This table uses the same specification as that in Column 6 of Table 7 and interacts "Friend" dummy with various measures on the extent of information asymmetry. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Rural to urban" indicates switchers who move from outside the city proper into the city proper. Six percent switchers move from the rural to urban part of the city. "Change sector" is one if the switcher changes the sector. "Call intensity" is the demeaned number of calls between switcher i and friends working at location j prior to the job switch. See Appendix Table A2 for summary statistics of key variables. In Columns 2-6, we also control for the baseline level of the interacted variable.

Table 12: The Referral Effect: Comparison with the Literature

Dependent variable Probability i switches to location j	(1)	(2)	(3)	(4)
Friend Definition				
Residential Neighbor	0.21***		0.18***	
	(0.01)		(0.01)	
Same Birth County		0.10***		0.09***
		(0.00)		(0.00)
Friend, not Neighbor			0.25***	
			(0.01)	
Friend, not Same Birth County				0.35***
				(0.02)
Controls	Yes	Yes	Yes	Yes
Observations	1,120,797	1,120,797	1,120,797	1,120,797
R-squared	0.16	0.11	0.20	0.16
New work Cell FE	No	No	No	No
OldxNew Cell FE	Yes	Yes	Yes	Yes
Num. of Cell FE	20,811	20,811	20,811	20,811

Notes: This table uses the same specification as that in Column 6 of Table 7. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Residential Neighbor" is a dummy that equals one if there is at least one individual who works in the new work location and shares the same residential location as the job switcher. "Same Birth County" is a dummy that equals one if there is at least one individual who works in the new work location and shares the same birth county as switcher i.

Table 13: Attributes of Referrals and Referees via a Dyadic Regression

Dependent variable 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
Num. of Cell FE	1,176	941

Notes: One unit of observation is a switcher-friend pair. A denotes the referred person and B denotes the referral. The dependent variable mean is 0.14. The sample restricts to switchers who eventually switch to some friend's workplace. Column 2 further restricts to switchers facing at least one vacancy in the same occupation that are posted in alternative locations in the same cell. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 14: Referral Benefits to Workers

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dependent variable	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: The 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 sizes among firms in the new work location. "ΔCoworker HP" is coworkers' average house price (thousand RMB) in the new workplace minus that in the 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 (with the majority of employees working in SOE firms), while the previous job is not. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. See Appendix Table A2 for summary statistics of key variables.

Table 15: Referral Benefits to Large Firms with Positive Hirings

Dependent variable				
Panel A: Log of Net Inflow	(1)	(2)	(3)	(4)
Referral	0.52***	0.51***	0.51***	0.53***
	(0.11)	(0.12)	(0.12)	(0.12)
Semi-elasticities	[0.71]	[0.70]	[0.71]	[0.73]
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.57	0.58	0.58	0.59
Panel B: Matching Rate	(5)	(6)	(7)	(8)
Referral	0.91***	0.88***	0.88***	0.86***
	(0.24)	(0.26)	(0.26)	(0.27)
Observations	[400,1000]	[400,1000]	[400,1000]	[400,1000]
R-squared	0.85	0.87	0.87	0.87
Panel C: Firm Growth Rate	(9)	(10)	(11)	(12)
Referral	0.49***	0.45***	0.44***	0.45***
	(0.11)	(0.10)	(0.10)	(0.11)
Observations	[600,1000]	[600,1000]	[600,1000]	[600,1000]
R-squared	0.76	0.83	0.83	0.83
Controls				
Firm Attributes	No	Yes	Yes	Yes
Previous Growth Rate	No	No	Yes	Yes
Employee Attributes	No	No	No	Yes
Cell FE	Yes	Yes	Yes	Yes

Notes: One unit of observation is a location with at least one matched firm that has more than 100 employees and positive hirings. Standard errors in round brackets. There are 239 cell fixed effects in Panel A, 190 in Panel B, and 271 in Panel C. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. "Referral" takes value 1 if there is at least one switcher moving to a friend in the firm. "Net inflow" is the number of switchers moving in minus moving out. "Log of net inflow" is log(Y+1) to include observations with zero values. In Panel A, "semi-elasticities" are reported in square brackets and are evaluated at the dependent variable mean. "Matching rate" is defined as the inflow over the number of vacancies. "Firm growth rate" is measured as the inflow over the employee size. Firm attributes include age, employee size, SOE dummy, and average real capital from 2010 to 2015. Previous firm growth rate is the average annual employee growth rate from 2010 to 2015. Employee attributes includes share of female, share of migrants, and the average age of pre-existing employees. Firm network size, measured by the number of distinct contacts of the firm's pre-existing employees, as well as the number of carrier A's users at each location is controlled in all columns. See Appendix Table A2 for summary statistics of key variables.

A Online Appendix

A.1 Occupancy Description

We use job descriptions and job titles and the US 2010 occupation codes to classify the occupation for each posting. Here are the occupations that we use:

1. Management – includes 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 – includes 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 – includes 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 – includes therapists, nurses, pharmacists, rehabilitation doctors, and 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, maintenance workers, gardeners, agricultural workers, forest workers, breeding workers, and livestock cultivators.

We combine the three smallest categories (Health related, Production and transportation, and Farming, fishing, and construction) into 'other category' in our empirical analysis.

A.2 Tables in the Appendix

Table A1: Summary Statistics of Diversity Measures

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

Notes: Each entropy measure is the normalized Shannon entropy averaged across either the working or the residential population at a given location. There are 6,161 locations in total.

Table A2: Summary Statistics of Key Variables in Regression Samples

Panel A: Switcher Attributes				
	Mean	SD	Median	N
Pr(i switches to j)	0.09	0.16	0.00	33,399
Friend	0.26	0.44	0.00	33,399
Distance(job1, job2) in km	10.45	15.72	3.95	38,102
Distance(home, job2) in km	8.58	12.95	3.29	34,927
Rural to urban	0.06	0.24	0.00	38,102
Young (Age 25-34)	0.36	0.48	0.00	38,102
Change sector	0.61	0.49	1.00	10,116
Panel B: Job Benefits				
	Mean	SD	Median	N
Wage at new job (thousand RMB)	31.47	24.30	25.22	17,615
ΔCoworker HP (thousand RMB)	-0.11	3.40	-0.06	23,323
PT to FT	0.16	0.37	0.00	19,431
Closer to home	0.31	0.46	0.00	29,117
Non-SOE to SOE	0.09	0.29	0.00	15,881
Panel C: Large firms with Positive hirings				
	Mean	SD	Median	N
Net inflow	2.77	6.35	1.00	[600,1000]
Matching rate	0.76	0.38	1.00	[600,1000]
Growth rate	0.04	0.06	0.02	[600,1000]
Firm network size (log)	5.92	1.90	6.13	[600,1000]
Referral	0.57	0.50	1.00	[600,1000]

Notes: Panel A reports summary statistics for key variables in Table 11. Panel B reports summary statistics for key variables in Table 14. Panel C reports summary statistics for key variables in Table 15.

Table A3: Referral Benefits to All Firms with Positive Hirings

	(1)	(2)	(3)
Dependent variable	Net inflow	Matching rate	Growth rate
Referral	0.36***	0.57***	0.49***
	(0.04)	(0.11)	(0.05)
Semi-elasticity	[0.50]	[0.57]	[0.49]
Controls	Yes	Yes	Yes
Observations	[3000,5000]	[2000,5000]	[3000,5000]
R-squared	0.44	0.79	0.70
Cell FE	Yes	Yes	Yes
Num. of Cell FE	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. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Referral Effects with an Alternative Friend Definition

Dependent variable						
Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.36***	0.36***	0.34***	0.34***	0.35***	0.35***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)	(0.01)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,151,676	1,120,797	1,151,676	1,120,797	1,151,676	1,120,797
R-squared	0.15	0.15	0.20	0.20	0.21	0.21
New work Cell FE	No	No	Yes	Yes	No	No
Old x New Cell FE	No	No	No	No	Yes	Yes
N. Cell FE	NA	NA	1,111	1,107	21,250	20,811

Notes: Same specification as in Table 7. Friends have at least four weeks' of non-missing work locations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Referral Benefits to Workers with an Alternative Friend Definition

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dependent variable	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 specifications as in Table 14. Friends have at least four weeks' of non-missing work locations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Referral Benefits to All Firms with an Alternative Friend Definition

	(1)	(2)	(3)
Dependent variable	Net inflow	Matching rate	Growth rate
Referral	0.66***	0.73***	0.62***
	(0.11)	(0.27)	(0.09)
Semi-elasticity	[0.90]	[0.73]	[0.62]
Controls	Yes	Yes	Yes
Observations	[600,1000]	[400,1000]	[600,1000]
R-squared	0.62	0.87	0.85
Cell FE	Yes	Yes	Yes
Num. of Cell FE	239	190	271

Notes: Same specifications as in Table 15. One unit of observation is a location with at least one matched firm and positive hirings. Friends have at least four weeks' of non-missing work locations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Referral Effect with a Two-way Friend Definition

Dependent variable						
Probability i switches to location j	(1)	(2)	(3)	(4)	(5)	(6)
Friend	0.39***	0.39***	0.37***	0.37***	0.38***	0.38***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)	(0.02)
Controls	No	Yes	No	Yes	No	Yes
Observations	1,151,676	1,120,797	1,151,676	1,120,797	1,151,676	1,120,797
R-squared	0.08	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	1,111	1,107	21,250	20,811

Notes: Same specifications as in Table 7. Friends are those who both place a call to and receive a call from individual i, or social contacts with two-way communications. Friends have at least four weeks' of non-missing work locations. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.