Online Appendix
"Information, Mobile Communication, and Referral Effects"

Panle Jia Barwick Yanyan Liu Eleonora Patacchini Qi Wu

S1 Additional Data Sets

Vacancy Data To gauge the dynamics of local labor market conditions, we collect listings from the two largest online job posting websites, Zhilian (zhaopin.com) and 58.com, from August 2016 to February 2018.[1] These websites hold on average 10,000 job postings per month. We obtained a total of 121,055 postings and merged them with our call data based on locations.

Each posting reports the posting date, job title and description, full time or part time, qualifications (minimum education level and years of experience), monthly salary (in a range), firm address, firm size (number of total employees), and firm industry. On the basis of the job title and description, we group these postings into eight occupations using the 2010 U.S. occupation code (see Online Appendix S2 for more details). Popular occupations include Professionals (26.70 percent), Service (26.61 percent), Sales and Office Administration (19.24 percent), and Management (17.47 percent), followed by Education, Legal, Arts and Media (11.53 percent), Farming, Fishing, and Construction (6.44 percent), Production and Transportation (2.29 percent), and Health Related (1.45 percent).

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

Administrative Firm-Level Records We use two firm-level administrative datasets to obtain wages and benefits, local industry composition, and firm attributes. The first is the annual National Enterprise Income Tax Records from 2010 to 2015, which is collected by

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

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

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

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

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

S2 Occupations on Job Postings

We use the job descriptions and job titles in the vacancy postings and the US 2010 occupation code to classify the occupation for each posting. The occupations we use are:

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, and nutritionists.

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

S3 Entropy Measures

The social entropy measures the diversity of individual i'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 $\frac{\nu_{ij}}{V_i}$, where ν_{ij} is the number of calls between i and j and V_i is the total number of calls placed or received by i. The denominator, log number of i's friends, is a scaling number that normalizes the Shannon entropy. Normalized entropy measures are guaranteed to vary between zero and one and are comparable across different measures, with higher values representing more diverse outcomes.

Inspired by the entropy literature, we propose to define income entropy as:

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

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

S4 Attributes of Referrers and Referees

To examine the characteristics of workers who find a job through referrals and those 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 of both referrer and referee attributes:

$$M_{ij} = \boldsymbol{X_i \alpha} + \boldsymbol{X_j \beta} + \boldsymbol{X_{ij} \gamma} + \lambda_c + \varepsilon_{ij}$$

where M_{ij} is one if i moves to friend j's workplace. The set of regressors includes $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$, which are gender, age, and birth county dummies for switcher i and friend j, and $\boldsymbol{X_{ij}}$, which includes dummies for whether i and j have the same gender and birth county, and an absolute difference in their age.

We limit the regression sample to the subset of switchers who find a job at some friend's workplace. The parameters are estimated via differences between dyads $\{i, j\}$ wherein i moves to j's work location and dyads $\{i, m\}$ where i does not move to m's work location. Column 1 of Table S7 includes all eligible dyads that have nonmissing demographic information, for a total of 93,000 observations. Column 2 only includes switchers for whom there is at least another location within the same neighborhood that has vacancy listings in the same occupation and salary range as the one that the job switcher takes. Females and migrant workers are more likely to receive referrals. Referral provision exhibits assortative patterns. 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 community networks based on birth county facilitate entry and the growth of private enterprises in China (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.

S5 Identify Employment Gaps

Identifying unemployment spells is challenging. The main issue is that changes in work locations (absence of a recurrent work location) may be due to factors different from unemployment such as part-time jobs, travel, sick leave, etc. In addition, as it can take time to find a job from unemployment, our analysis of the referral effect at the time of reemployment is necessarily limited to employed individuals who experience unemployment and find another (stable) job within twelve months.

Our strategy to minimize measurement errors is to use stringent sample requirements and add robustness checks using different definitions. We first select individuals with a stable job for at least four consecutive weeks (about 75 percent of the sample). We then limit the sample to individuals who experience an unemployment spell (a period without recurring work locations) for at least eight weeks and exclude individuals returning to the old work location after the unemployment spell. We require the maximum hours in any location during unemployment to be fewer than 20 hours. These requirements leave us with a total of 21,542 individuals. Among these individuals, our final sample consists of those who find another stable job during the sample period, that is, those with another recurrent job location (different from the previous job location) for at least four weeks. We end up with 5,164 job switchers who experience an employment gap (reemployed workers), of whom 3,638 have valid friend work locations and 1,677 find the new job through referrals.

Using these definitions, the share of reemployed workers in our data ($0.24 = 5{,}164/21{,}542$) is similar to the official measure of the reemployment rate in China, as reported by the China Labor-force Dynamics Survey (CLDS) for the year 2015 (the year before our call data were collected). In this survey, respondents between 25 and 60 years old are asked about their unemployment history. Among all respondents who were unemployed in 2015, 24 percent reported successfully finding a new job in the same year. While we use a different sample from the CLDS, our number on the reemployment rate is remarkably in line with those reported by CLDS. In addition, the share of reemployed workers who found jobs through referrals (between 1,677/5,164=32.5 percent and 1,677/3,638=46 percent) is similar to the share of individuals who reported finding jobs through friends in the 2014 China Family Panel Studies survey (38 percent).

The sample for event study Figure 5 and regression Table S21 uses a ceiling of 20 hours at any location for the entire duration of unemployment (the average hours worked is 12). We repeat the event study and referral regressions with different thresholds (such as a maximum of 15 or 10 hours at any location and longer or shorter unemployment spans) and obtain qualitatively similar findings. Figure S3 reports event studies using alternative thresholds for unemployment. The patterns are similar to Figure 5 in the main text.

References

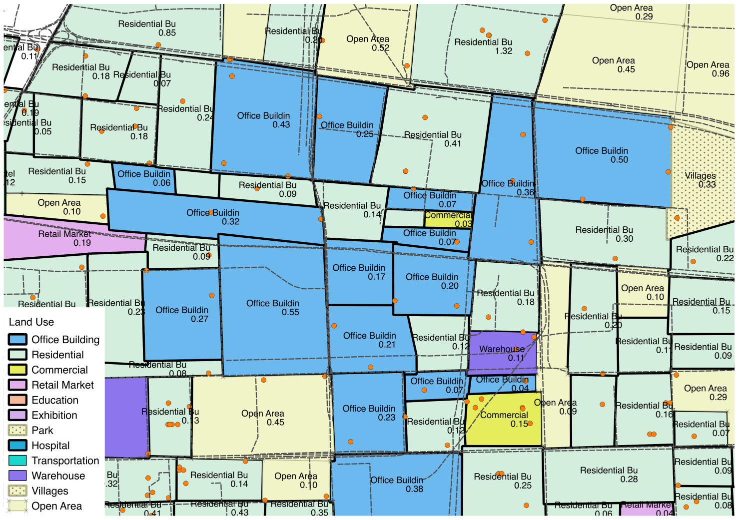
Cai, Hongbin, Hanming Fang, and Lixin Collin Xu. 2011. "Eat, Drink, Firms, Government: An Investigation of Corruption from the Entertainment and Travel Costs of Chinese Firms." *The Journal of Law and Economics*, 54(1): 55–78.

Dai, Ruochen, Dilip Mookherjee, Kaivan Munshi, and Xiaobo Zhang. 2018. "Community networks and the growth of private enterprise in China." *VoxDev.*

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

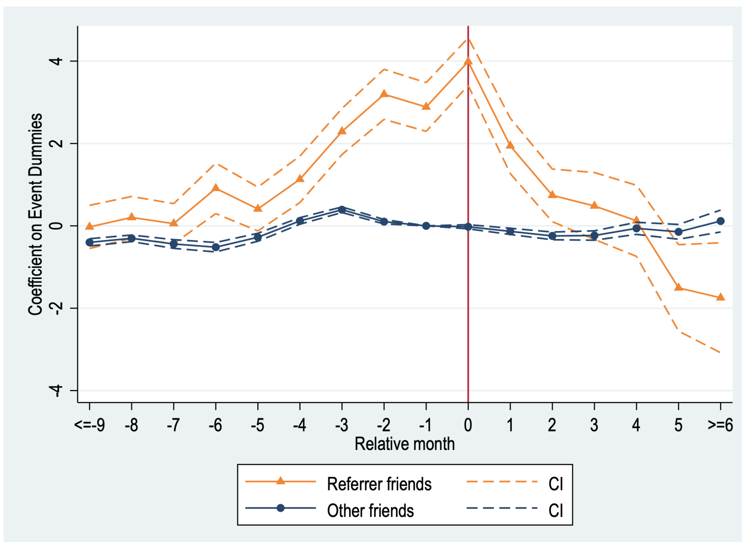
S6 Additional Figures and Tables

Figure S1: Neighborhoods and Locations



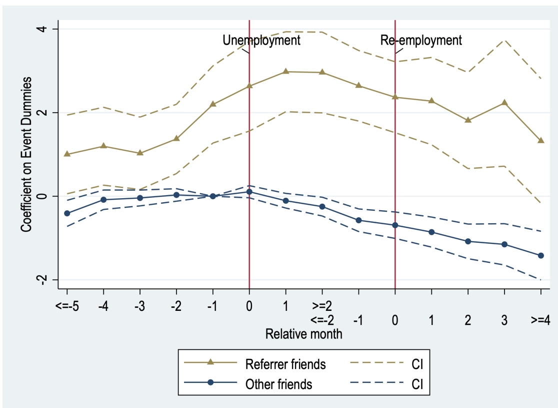
Notes: This map illustrates neighborhoods (polygons separated by dark lines) and locations (orange dots) as well as their corresponding land use (denoted by different colors and shades) in a section of the city we study. The city is divided into 1,406 neighborhoods that are delineated by major roads, with a total of 17,881 locations. The number in each polygon denotes the area size in km^2.

Figure S2: Calls to Referrer and Nonreferrer Friends, Controlling for Initial Call Intensity

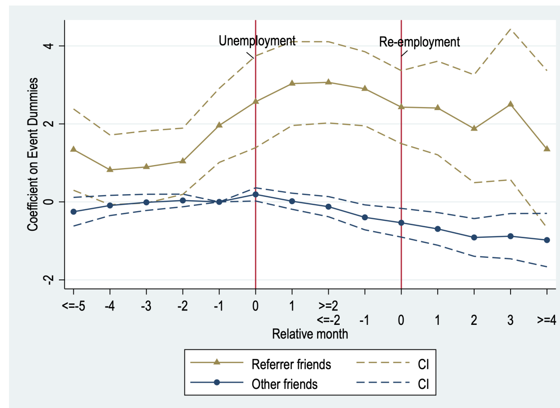


Notes: This graph replicates the baseline event study Figure 3 but controls for tie strength, proxied by the initial communication intensity between switchers and friends. Specifically, we control for the number of calls between switcher i and friend j during the first three months as a fraction of i's total calls during the same period. The orange line (with triangles) represents calls between switchers and their referrals while the blue line (with dots) represents calls between switchers and their nonreferrer friends. The vertical line indicates the month of job switch. The reference group is the call frequency between switchers and nonreferrer friends one month prior to the job switch. Switcher fixed effects and calendar month fixed effects are included in the regression. We thank one of the referees for suggesting this event study.

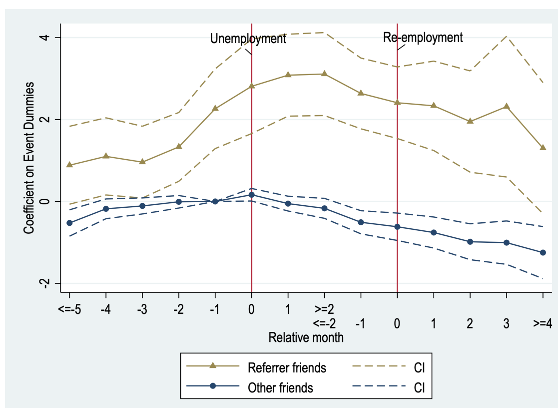
Figure S3: Calls to Referrer and Nonreferrer Friends by Individuals with Unemployment Spells: Robustness



(a) Max 15 hours



(b) Max 15 hours, individuals between 25 and 60



(c) Max 15 hours, 8+ weeks at the old job

Notes: Similar to Figure 5, this figure plots the coefficient estimates and their 95% confidence intervals for an event study that examines the number of calls between a reemployed individual with unemployment spells and his/her referrer and nonreferrer friends using different unemployment definitions. The first vertical line denotes the month immediately before unemployment. The second vertical line denotes the month of reemployment. The brown lines (with triangles) denote calls between reemployed individuals and their referrer friends. The blue lines (with dots) denote calls between reemployed individuals and their nonreferrer friends. In Panel (a), the maximum number of hours at any location during unemployment is capped at 15 hours. Panel (b) is the same as Panel (a), except that it limits to individuals between 25 and 60 years old. Panel (c) further restricts to individuals with at least 8 weeks in the previous workplace.

Table S1: Percentage of Job Switchers Moving to a Friend's Workplace

Panel A: friends with at least 45 weeks of location information			
	Percent	Number of individuals	Number of dyads
Switching to a friend's place	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	Number of individuals	Number of dyads
Switching to a friend's place	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 Section 2.1. Panel A includes all friends with nonmissing work locations for at least 45 weeks. Panel B includes all friends with nonmissing work locations for at least 4 weeks. "Switching to a friend" takes value one if a switcher moves to a preexisting friend's workplace. "Missing all friends' locations" reports the number of switchers with no valid location information for any preexisting friend. "Number of dyads" is the number of switcher-friend pairs.

Table S2: Job Changes and Information Flow Between Districts

	(1) Share of job changes	(2) Share of calls
Between urban districts	49.59%	63.68%
Between rural districts	34.99%	13.08%
Between urban and rural districts	15.42%	23.24%
Between high-income districts	73.48%	80.91%
Between low-income districts	18.94%	13.11%
Between high and low-income districts	7.58%	5.98%
Between high-amenity districts	75.35%	82.08%
Between low-amenity districts	17.61%	12.26%
Between high and low-amenity districts	7.05%	5.67%

Notes: This table shows the spatial patterns of job changes and phone communications between administrative districts. Column 1 shows the fraction of job switchers changing jobs between district pairs of different levels of socioeconomic development. Column 2 shows the fraction of calls between different district pairs. "Urban" districts refer to the urban core of the city we study, while "rural" districts include all other districts and surrounding counties. A district is "high/low-income" if the district's average house price is above/below the median of the distribution across all districts. The number of amenities is the total number of restaurants, major roads and parking lots, and schools in each district. A district is "high/low-amenity" if the district's average number of amenities is above/below the median of the distribution across all districts.

Table S3: Information Flow and Worker Flows

Dependent variable:	District pairs		Neighborhood pairs		Location pairs	
Worker flows (l, k)	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Regression analysis					
Information flow (l, k) (thousand)	n.a.	3.25***	n.a.	0.10***	n.a.	0.05***
		(0.39)		(0.02)		(0.01)
Observations	251	251	987,713	987,713	159,856,138	159,856,138
R-squared	0.745	0.971	0.024	0.164	0.002	0.042
Area l + Area k	Yes	Yes	Yes	Yes	Yes	Yes
	Panel B: Prediction exercise					
RMSE	66.610	26.251	0.366	0.178	0.012	0.011
MAPE	596.736%	115.267%	2.610%	1.011%	0.0099%	0.0098%

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

Table S4: Summary Statistics of Key Variables in Regression Samples

Panel A: switcher attributes	Mean	SD	N
Pr(i switches to l)	0.09	0.16	33,399
Friend	0.26	0.44	33,399
Distance(job1, job2) in km	10.45	15.72	38,102
Distance(home, job2) in km	8.58	12.95	34,927
Rural to urban	0.06	0.24	38,102
Young (Age 25-34)	0.36	0.48	38,102
Change sector	0.61	0.49	10,116
Panel B: job benefits	Mean	SD	N
Wage at new job (thousand RMB)	31.47	24.30	17,615
ΔCoworker HP (thousand RMB/m^2)	-0.11	3.40	23,323
PT to FT	0.16	0.37	19,431
Shorter commute	0.31	0.46	29,117
Non-SOE to SOE	0.09	0.29	15,881
Panel C: firm attributes	Mean	SD	N
Net inflow	2.77	6.35	[600,1000]
Matching rate	1.53	2.33	[600,1000]
Growth rate	0.04	0.06	[600,1000]
Firm network size (log)	5.92	1.90	[600,1000]
Referral	0.57	0.50	[600,1000]

Notes: Panel A reports summary statistics for key variables in Table 4. Panel B reports summary statistics for key variables in Table 7. Panel C reports summary statistics for key variables in Table 8.

Table S5: Referral Effect with Progressively More Controls

Dependent variable: Probability i switches to location l	(1)	(2)	(3)
Friend	0.36***	0.36***	0.34***
	(0.003)	(0.003)	(0.01)
Observations	1,151,676	1,120,797	1,120,797
R-squared	0.08	0.08	0.14
Controls	No	Yes	Yes
Old x new work neighborhood FE	No	No	Yes
Number of neighborhood-pair FE	NA	NA	20,811

Notes: This table uses the same specification and controls as those in Column 1 of Table 3, with an increasingly more saturated set of controls. The unit of observation is a switcher-location pair. "Friend" is a dummy variable that equals one if individual i has at least one friend working at a given location. Column 1 has no controls or fixed effects. Column 2 includes location attributes as well as interactions between location and demographic attributes. Column 3 includes neighborhood-pair fixed effects. Standard errors are clustered by neighborhood pair in Column 3 and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S6: The Effect of Friends of Friends: Additional Evidence

Dependent variable: Probability i switches to location l	(1)	(2)	(3)
Friend	0.34***	0.37***	0.32***
	(0.01)	(0.04)	(0.04)
Friend of i's nonreferrer friends	0.14***	0.11***	0.09***
	(0.01)	(0.03)	(0.02)
Old occupation	No	Yes	Yes
Number of calls at location l	No	No	Yes
Observations	915,251	915,251	915,251
R-squared	0.126	0.128	0.134
Controls	Yes	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes
Number of neighborhood-pair FE	16468	16468	16468

Notes: This table examines the coefficient of friends of friends (Column 5 in Table 3) and shows that it is partly driven by occupation clusters in addition to homophily. Column 1 replicates column 5 in Table 3. Column 2 controls for the old occupation of the switcher, and interacts the occupation dummies with "Friend" and "Friend of i's nonreferrer friends". The reference occupation is "retail workers". The interactions between friends of friends and occupation dummies are only significantly different from the default group (nonreferrer friend in retail) for finance and professional services. The coefficients for "finance × friends of friends" and "professional service × friends of friends" are 0.18 and 0.1 higher than the reference group, respectively. Similarly, the interactions between "Friend" and occupation dummies are significantly different from the default group (nonreferrer friends in retail) for finance, professional service, and wholesale, whose coefficients are 0.16 and 0.12 higher than and 0.16 lower than the default group, respectively. Column 3 further controls for the de-meaned number of calls associated with location l (excluding the switcher's calls) and the interactions with "Friend" and "Friend of i's nonreferrer friends" to capture job clusters in high-density locations. The reduction in the coefficient for "Friend of i's nonreferrer friends" across columns suggests that it is partially driven by job clusters among finance and professional services. Nonetheless, referral effect is robust to controlling for occupation clusters. Standard errors are clustered by the neighborhood pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S7: Attributes of Referrals and Referees via a Dyadic Regression

Dependent variable: Probability that A switches to B's workplace	(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
New work neighborhood FE	Yes	Yes
Number of neighborhood FE	1,176	941

Notes: This table examines the characteristics of potential referrer-referee pairs. The dependent variable takes value one for referrer-referee pairs. The sample restricts to switchers who eventually move to a referrer's workplace and includes all friends of these switchers. The unit of observation is a switcher-friend pair. "A" denotes the job switcher and "B" denotes the referrer. The dependent variable mean is 0.14. Column 2 limits to switchers facing at least one vacancy in alternative locations of the same neighborhood that is in the same occupation and salary range as the one switchers take. Standard errors are clustered by neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S8: Number of Calls between Residential Neighbors, Co-workers, and People with Same Birth County

Dependent variable: Number of calls between (i,j) per month	(1) Random 1% individuals living in same neighborhood	(2) Random 1% individuals working in same neighborhood	(3) Random 1% individuals working in same neighborhood
Same residential location	0.009***		
	(0.002)		
Same work location		0.012***	
		(0.003)	
Same birth county			0.002**
			(0.001)
Observations	1,801,076	1,130,498	1,130,498
R-squared	0.034	0.004	0.004
Dependent variable mean	0.002	0.003	0.003
Residential neighborhood FE	Yes	No	No
Work neighborhood FE	No	Yes	Yes
Number of neighborhood FE	1051	1012	1012

Notes: This table validates the literature's traditional measures of social interaction by examining the communication patterns. One unit of observation is a pair of individuals (i,j). The dependent variable is the average number of calls between i and j per month (its sample mean is 0.002, 0.003, and 0.003 in each column, respectively). Column 1 uses pairs of individuals from a 1 percent random sample of residents in each neighborhood. The dummy variable "Same residential location" indicates whether i and j live in the same residential location (a smaller geographical unit than a neighborhood). Column 2 uses pairs of individuals from a 1 percent random sample of workers in each neighborhood. The dummy variable "Same work location" indicates whether i and j work in the same location. Column 3 uses the same sample as in Column 2. "Same birth county" takes value one if i and j are born in the same county. Column 1 controls for residential neighborhood fixed effects. Columns 2 and 3 control for work neighborhood fixed effects. Standard errors are clustered at the same level as the fixed effects and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S9: Referral Effect and Tie Strength

Dependent variable: Probability i switches to location l	(1)	(2)
Friend	0.34***	0.35***
	(0.01)	(0.01)
Friend× Call intensity	0.0002***	
	(4.20e-05)	
Friend× $\text{Call}_{il}/\text{Call}_i$		0.38***
		(0.02)
Observations	915,251	915,251
R-squared	0.13	0.14
Old x new work neighborhood FE	Yes	Yes
Number of neighborhood-pair FE	16,468	16,468

Notes: This table uses the same specification as that in Column 2 of Table 3 and interacts the "Friend" dummy with measures of tie strength between the referrer pair. "Call intensity" is the de-meaned number of calls between switcher i and referrer friend l prior to the job switch. $\text{Call}_{il}/\text{Call}_i$ is the de-meaned ratio – the number of calls between i and l as a fraction of i's total number of calls prior to the job switch. Standard errors are clustered by the neighborhood pair and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table S10: Social Contact Diversity and Probability of Using Referrals

Dependent variable: Probability of job change using referrals	(1)	(2)	(3)	(4)
Social entropy		0.102***		0.074***
		(0.016)		(0.017)
Income entropy			0.039***	0.022**
			(0.009)	(0.010)
Number of calls (thousand)	0.020***	0.020***	0.019***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	34,812	34,812	34,812	34,812
R-squared	0.089	0.090	0.090	0.091
Residential neighborhood FE	Yes	Yes	Yes	Yes
Number of neighborhood FE	1,118	1,118	1,118	1,118

Notes: This table examines whether individuals with more diverse social contacts are more likely to use referrals. The unit of observation is a job switcher. "Probability of job change using referrals" is one if the individual moved to a new workplace with at least one friend, zero otherwise. Social entropy and income entropy are normalized Shannon entropies as defined in Online Appendix Section S3. Number of calls (thousand) is the total number of calls in thousands that originate from or are received by an individual. We control for demographics (age, gender, birth city and county) and the fraction of strong ties in all columns (results are robust whether we use the 75th or 90th percentile to define strong ties). Standard errors are clustered by neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S11: Referral Effect with Different Restrictions on Preexisting Contacts

Dependent variable: Probability i switches to location l	(1)	(2)	(3)	(4)	(5)
Excluding links formed within 1 to 5 months of the job switch					
Friend (1 months)	0.35***				
	(0.01)				
Friend (2 months)		0.35***			
		(0.01)			
Friend (3 months)			0.35***		
			(0.01)		
Friend (4 months)				0.34***	
				(0.02)	
Friend (5 months)					0.34***
					(0.02)
Observations	915,251	915,251	915,251	915,251	915,251
R-squared	0.135	0.128	0.124	0.100	0.093
Controls	Yes	Yes	Yes	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes	Yes	Yes
Number of neighborhood-pair FE	16,468	16,468	16,468	16,468	16,468

Notes: This table examines changes in the referral effect estimate when using different definitions of preexisting social contacts. The table uses the same specification as that in Column 2 of Table 3, but an increasingly more stringent cutoff for referrer friends. For example, Column 1 excludes friends formed within 1 month of the job switch, while Column 5 excludes all friends formed within 5 months of the job switch. Standard errors are clustered by the neighborhood pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S12: Referral Effect with Alternative Friend Definitions

Dependent variable: Probability i switches to location l	(1) Four weeks	(2) Two-way	(3) Above median	(4) Below median
Friend	0.36***	0.38***	0.34***	0.36***
	(0.01)	(0.02)	(0.02)	(0.01)
Observations	915,251	915,251	496,825	418,426
R-squared	0.19	0.13	0.14	0.11
Controls	Yes	Yes	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes	Yes
Number of neighborhood-pair FE	16,468	16,468	9,335	8,546

Notes: This table examines the referral effect's robustness to alternative friend definitions using the baseline specification (Column 2 in Table 3). Column 1 includes all friends with at least four weeks of nonmissing work locations. In Column 2, friends are social contacts with two-way communications (they both place calls to and receive calls from switcher i) and at least four weeks of nonmissing work locations are included. As we do not observe locations for friends outside Company A's subscriber network, Columns 3 and 4 split the switcher sample based on whether the fraction of a switcher's social contacts in Company A is above or below the median (the cutoff is 48 percent). Standard errors are clustered by the neighborhood pair and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S13: Referral Benefits for Workers with Alternative Friend Definition

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Wage at new job	ΔCoworker HP	PT to FT	Shorter Commute	Non-SOE to SOE
Friend	0.40*	0.08**	0.02***	0.09***	0.0075*
	(0.21)	(0.04)	(0.01)	(0.01)	(0.004)
Observations	18,595	24,835	21,016	31,013	16,789
R-squared	0.79	0.52	0.10	0.12	0.56
Residence neighborhood FE	Yes	Yes	Yes	Yes	Yes
New work neighborhood FE	Yes	Yes	Yes	Yes	Yes

Notes: This table uses the same specification as those in Table 7 but includes all friends with at least four weeks of nonmissing work locations. Standard errors are two-way clustered by the residential neighborhood and new work neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S14: Referral Benefits for Workers and Other Types of Friends

	Income Effect		Job Quality		
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Wage at new job	ΔCoworker HP	PT to FT	Shorter Commute	Non-SOE to SOE
Panel A: baseline (Table 7)					
Friend	0.62**	0.07*	0.014**	0.09***	0.012**
	(0.31)	(0.04)	(0.007)	(0.01)	(0.005)
Panel B: with proxies for other types of friends					
Friend	0.62***	0.07*	0.02**	0.09***	0.012***
	(0.22)	(0.04)	(0.01)	(0.01)	(0.005)
Observations	17,615	23,323	19,431	29,117	15,881
Residence neighborhood FE	Yes	Yes	Yes	Yes	Yes
New work neighborhood FE	Yes	Yes	Yes	Yes	Yes

Notes: Panel A presents results in Table 7. Panel B uses the same specification as those in Table 7, but also controls for"Friend who moved before the job change", "Friend living but not working in i's new workplace" and "Friend of i's nonreferrer friends". Standard errors are two-way clustered by the residential neighborhood and new work neighborhood and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table S15: Referral Benefits for All Firms with Positive Hiring

	(1)	(2)	(3)
Dependent variable:	Net inflow	Matching rate	Growth rate
Referral	0.46***	0.57***	0.49***
	(0.05)	(0.11)	(0.05)
Observations	[3000,5000]	[3000,5000]	[3000,5000]
R-squared	0.53	0.79	0.70
Controls	Yes	Yes	Yes
Neighborhood FE	Yes	Yes	Yes
Number of Neighborhood FE	631	526	707

Notes: This table uses the same specification as that in Column 4 of Table 8 but includes all firms with positive hiring. Standard errors are reported in parentheses and clustered by neighborhood. *** p<0.01, ** p<0.05, * p<0.1.

Table S16: Referral Benefits for Large Firms with Alternative Friend Definition

Dependent variable:	(1) Net inflow	(2) Matching rate	(3) Growth rate
Referral	0.81***	0.73***	0.62***
	(0.13)	(0.27)	(0.09)
Observations	[600,1000]	[600,1000]	[600,1000]
R-squared	0.68	0.87	0.85
Controls	Yes	Yes	Yes
Neighborhood FE	Yes	Yes	Yes
Number of Neighborhood FE	225	190	271

Notes: This table uses the same specification as that in Column 4 of Table 8 but includes friends with at least four weeks of nonmissing work locations. Standard errors are reported in parentheses and clustered by neighborhood. *** p<0.01, ** p<0.05, * p<0.1.

Table S17: Phone Calls, Internet Browsing, and Data Usage

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

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

Table S18: Phone Calls, Text Messages, and WeChat

Dependent variable: Panel A: Number of calls	(1)	(2)	(3)	(4)	(5)
Number of text messages	11.48***				
	(3.46)				
WeChat voice traffic (GB)		4.17***	1.63***		
		(0.69)	(0.32)		
Total WeChat traffic (GB)				6.65***	2.19***
				(0.77)	(0.32)
Dependent variable: Panel B: Minutes of calls	(6)	(7)	(8)	(9)	(10)
Number of text messages	10.12***				
	(3.18)				
WeChat voice traffic (GB)		6.58***	3.52***		
		(0.91)	(0.65)		
Total WeChat traffic (GB)				9.57***	4.11***
				(1.00)	(0.63)
Observations	20,000	79,912	79,912	79,912	79,912
Caller FE	No	No	Yes	No	Yes
Week FE	No	Yes	Yes	Yes	Yes

Notes: This table examines whether different communication channels (phone calls, text messages, and apps such as WeChat) are complements using phone activities of 20,000 randomly selected cellphone subscribers. Data on WeChat usage are available for four weeks from October 29 to November 25, 2020. Data on text messages are available for the week of November 19 to November 25, 2020. Each cell in the table denotes a separate regression. The dependent variable is the number of weekly calls in Panel (a) and call duration in Panel (b). 'WeChat voice traffic" refers to the mobile data traffic that is associated with WeChat video and audio calls. "Total WeChat traffic" refers to the total data flow of WeChat activities including messages, calls, files sharing, etc. Columns 2-5 and 7-10 include week fixed effects. Columns 3,5,8 and 10 also include individual fixed effects. Standard errors in Columns 2-5 and 7-10 are clustered at the individual level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table S19: Referral Effect and Communication Technology: 4G Compatibility

Dependent variable: Probability i switches to location l	(1) Baseline	(2) 4G	(3) non-4G
Friend	0.35***	0.35***	0.34***
	(0.01)	(0.01)	(0.02)
Observations	915,251	700,326	214,925
R-squared	0.12	0.12	0.13
Controls	Yes	Yes	Yes
Old x new work neighborhood FE	Yes	Yes	Yes
Number of neighborhood-pair FE	16,468	13,387	4,289

Notes: This table examines the robustness of the estimated referral effect to 4G technology using the baseline specification (Column 2 in Table 3). Column 1 reports the baseline specification. Columns 2 and 3 repeat the baseline analysis separately for 4G and non-4G users, as individuals with access to a fast Internet (the 4G network) could use mobile apps (such as WeChat) instead of calls to communicate with their friends. Standard errors are clustered by neighborhood pair and reported in parentheses in all columns. *** p<0.01, ** p<0.05, * p<0.1.

Table S20: Communication with Friends During Job Search

Dependent variable: Average monthly calls during search period	(1)	(2)	(3)	(4)	(5)	(6)
Unemployed	40.10***	35.80***	34.31***	32.94***	30.64***	35.26***
	(2.02)	(1.14)	(1.16)	(1.44)	(2.05)	(2.07)
Observations	38,830	38,793	38,793	38,793	19,507	19,286
R-squared	0.01	0.50	0.51	0.53	0.54	0.54
Average calls outside the search period	No	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Residential neighborhood FE	No	No	No	Yes	Yes	Yes

Notes: This table examines whether switchers experiencing unemployment spells make more calls than on-the-job switchers. The dependent variable is the average monthly calls associated with all social contacts during the search period. The search period is the unemployment period for reemployed workers and the three months prior to the job switcher for on-the-job switchers. The regression sample consists of all switchers with and without unemployment spells. "Unemployed" is a dummy that takes value one for those with unemployment spells. Columns 2-6 control for the average monthly calls outside the search period. Columns 3-6 add individual demographics (gender, age group dummies, migrant) and Columns 4 to 6 further include residential neighborhood fixed effects. In Columns 5 and 6, we divide reemployed workers into two groups based on whether their unemployment length is below (Column 5) or above (Column 6) the median of 11 weeks. We randomly allocate on-the-job switchers into Columns 5 and 6 so the number of observations in these columns is comparable and sums up to the total number of observations in Column 4.

Table S21: Referral Effect for Individuals with Unemployment Spells

Dependent variable: Probability i reemployed at location l	(1)	(2)	(3)
Friend	0.33***	0.33***	0.33***
	(0.01)	(0.01)	(0.01)
Observations	166,924	166,888	166,888
R-squared	0.04	0.04	0.12
Controls	No	Yes	Yes
Old x new work neighborhood FE	No	No	Yes
Number of neighborhood-pair FE	NA	NA	4,268

Notes: This table examines the referral effect for workers who are laid off and then become reemployed during the sample period (those with unemployment spells). The unit of observation is a reemployed-worker and location pair. "Friend" is a dummy variable that equals one if individual i has at least one friend working at a given location. Column 1 has no controls or fixed effects. Column 2 includes location attributes and interactions between location and demographic attributes as in Column 1 of Table 3. Column 3 controls for the old-by-new work neighborhood-pair fixed effects. Standard errors are clustered by neighborhood pair and reported in parentheses in Column 3. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.