Information, Mobile Communication, and Referral Effects

Panle Jia Barwick, Yanyan Liu, Eleonora Patacchini, Qi Wu

NBER Working Paper

ACE 592

Sep 19, 2019

Outline

Outline

Introduction

- Motivation:
  - The effect of information exchange on economic outcomes
  - Difficult to measure
  - Unique data set: universe of cellphone usage records with geocoded information
- Research Questions:
  - Investigate the extent to which information flow is accompanied by worker flows
  - Examine how information flow among friends affects job transitions and worker-vacancy match efficiency

Introduction

- Literature Review:
    - Digital footprints in urban economics
      Henderson et al. (2012), Akbar et al. (2018)
    - Information economics and diversity
      Ottaviano and Peri (2006), Eagle et al.(2010), Ashraf and Galor (2011)
    - Social networks in job searches
      Ioannides and Loury (2004), Bayer et al. (2008)
- Findings:
    - Information flow (measured by call volume) correlates strongly with worker flow
    - Referral information is valuable to both workers and firms

Outline

1. Introduction

2. **Motivation and Data**

3. Estimation

4. Conclusion and Discussion

Data

- Call data:
    - Phone records for 1.6 million mobile-phone users in a northern city in China from November 2016 to October 2017
    - A major telecommunication operator Serves 30-65% of all mobile phone users with 85% cellphone penetration rate
- Vacancy data:
    - Top two largest online job posting websites
- Administrative Firm-Level Records:
    - National Enterprise Income Tax Records for economic variables
    - Business Registration Database for location data
- House price:
    - A major online real estate brokerage intermediary
- Key variables: Friend and Working/Residential location

Labor Market Summary Statistics

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.

Figure 1: Information Flow and Worker Flow Among Administrative Districts
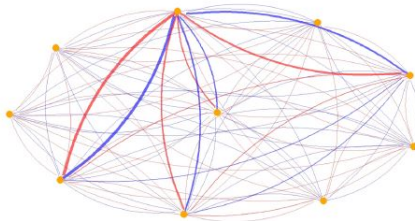
Notes: Each node is an administrative district in the city. We plot randomly selected ten nodes out of a total of 23. Blue (non-directional) edges correspond to the number of job switches among the pairs of nodes, with the width of each edge scaled proportionately to the number of switches. Red edges 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.

Table 1: Information Flow and Worker Flows

(a) At the Administrative District Level

Dependent variable: Worker flow (i, j)	All calls	Calls from/to job switchers ***before*** switch		
		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

Dependent variable: Worker flow (i, j)	All calls	Calls from/to job switchers ***before*** switch		
		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.

Diversity and Economic Outcome

- Social entropy

$$D^{Social}(i) = -\frac{\sum_j P_{ij} * log(P_{ij})}{log(NumFriend_i)}$$
$$= -\frac{\sum_j \frac{v_{ij}}{V_i} log(\frac{v_{ij}}{V_i})}{log(NumFriend_i)}$$

- Spatial entropy

$$D^{Spatial}(i) = -\frac{\sum_l P_{il} * log(P_{il})}{log(NumLocation_i)}$$

- Income entropy

$$D^{Income}(i) = -\frac{\sum_d P_{id} * log(P_{id})}{log(NumDecile_i)}$$

Diversity and Job Switches

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.

Diversity and Job Switches

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 entropy	-0.10		
	(0.28)		
Spatial entropy		-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

Outline

Event Study

- Time series variation of information exchange between job seekers and referral vs. non-referral friends
- A battery of regression analyses to illustrate that our estimated referral effect is not driven by confounding factors
- Restricted to network formed three month ago

Event Study

- The weekly number of contacts prior to and after the job change is stable

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.

Event Study

- To check the dynamic of information flow, the following relationship is examined for an event window of 11 months before and 10 months after the switch

$$Freq_{ijt} = c + \sum_{s=-11}^{10} \gamma_s Referral_{ij}[t = s] + \sum_{s=-11}^{10} b_s NonReferral_{ij}[t = s]$$
$$+ \lambda_i + \tau_t + \epsilon_{ijt}$$

- $Freq_{ijt}$ 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
- λ_i is an individual fixed effect
- τ_t is a month fixed effect

Event Study

Figure 6: Event Study – Number of Calls to Referrals vs. Non-referrals
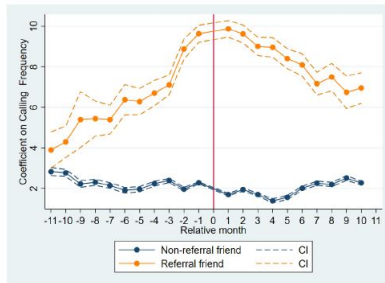
Notes: orange line represents calls between switchers and the referrals (Obs = 252,852). Blue line represents calls between switchers and non-referrals (Obs = 4,915,656). Switcher fixed effects and month fixed effects are included in the regression.

- Switchers have more frequent calls with referral friends
- The intensity of information flow between switchers and their referrals exhibits an inverted-U shape
- Communication intensity difference post job remains

Referrals and Work Choices

- To check the effect of referrals

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

- 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
- λ_c denotes location pair fixed effects
- X_{ki} denote demographic controls

Referrals and Work Choices

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).

- The mean propensity to choose a location among the switchers is 0.09
- The referral effect, which ranges from 0.34 to 0.36, is economically large and stable

Referral effect on labor and firm

- To check the effect on labor:

$$Y_{ilr} = \beta Friend_{ilr} + \sum_{k=1}^{K} \beta_k X_{ki} + \lambda_c + \alpha_r + \epsilon_{il}$$

  - 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

- Similarly for the effect on firm :

$$Y_i = \gamma Referral_i + \sum_{k=1}^{K} \beta_k Z_{ki} + \lambda_c + \epsilon_i$$

  - i denotes a firm
  - Y_i firm performance: net inflow of workers; match rate; and firm growth

Table 14: Referral Benefits to Workers

Dependent variable	Income Effect		Job Quality		
	(1) Wage at new job	(2) ΔCoworker HP	(3) PT to FT	(4) Closer to home	(5) 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.

- Referral jobs are associated with Wage premium, higher housing price, increase in the likelihood of working full-time, shorter commute and probability of moving to SOE job

Referral effect on firm

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

- Firms that grow quickly are more likely to hire through employee referrals, the estimate could be biased upward
- Results in Table 15 are largely descriptive

Outline

Coonclusion and Discussion

- Conclusion:
  - Information flow (measured by call volume) correlates strongly with worker flow
  - Referral information is valuable to labor market outcome and equality
  - Referral information is important for firm growth
- Contribution:
  - Mobile communication records is a better source of information flow than conventional data set, e.g. survey data
  - Combination of big data and traditional data sources
- Limitation:
  - Descriptive, mechanism of how information flow influences labor flow is still unclear
  - Suffers from endogenous problem in causal inference
- Alternative? Any construction on friend? Other idea from using call data?