DISCUSSION PAPER SERIES

IZA DP No. 8089

Strong versus Weak Ties in Migration

Corrado Giulietti
Jackline Wahba
Yves Zenou

April 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Strong versus Weak Ties in Migration

Corrado Giulietti
IZA

Jackline Wahba
University of Southampton,
CPC and IZA

Yves Zenou
Stockholm University, IFN,
University of Southampton and IZA

Discussion Paper No. 8089
April 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 8089
April 2014

ABSTRACT

Strong versus Weak Ties in Migration*

This paper studies the role of strong versus weak ties in the rural-to-urban migration decision in China. We first develop a network model that puts forward the different roles of weak and strong ties in helping workers to migrate to the city. We then use a unique longitudinal data that allows us to test our model by focusing on first-time migration. Strong ties are measured by the closest family contact, while weak ties are determined by the fraction of migrants from the village in which the individual resides. We address the endogeneity of the network formation in the migration decision. Our results indicate that both weak and strong ties matter in the migration decision process, although the impact of weak ties is higher than that of strong ties. We also show that one underestimates the effect of social networks on migration by not taking into account the strong ties in the mobility process. We finally find that weak and strong ties act as complements in the migration decision, which indicates that the interactive effect between weak and strong ties is particularly strong above a certain threshold of the size of weak ties.

NON-TECHNICAL SUMMARY

This paper explores for the first time the joint role of strong and weak ties in the migration decisions. Using data from rural-to-urban migration in China, we find that an individual is more likely to migrate if there are more migrants in his/her network. This holds for both strong ties (the closest contact) and weak ties (the village population). Remarkably, we find that there are complementarities between the two types of networks, in that the role of weak ties is stronger when a strong tie has migrated.

JEL Classification: O15, J61

Keywords: social networks, internal migration, China

Corresponding author:

Corrado Giulietti
IZA
P.O. Box 7240
53072 Bonn
Germany
E-mail: giulietti@iza.org

* The Longitudinal Survey on Rural Urban Migration in China (RUMiC) consists of three parts: the Urban Household Survey, the Rural Household Survey and the Migrant Household Survey. It was initiated by a group of researchers at the Australian National University, the University of Queensland and the Beijing Normal University and was supported by the Institute for the Study of Labor (IZA), which provides the Scientific Use Files. The financial support for RUMiC was obtained from the Australian Research Council, the Australian Agency for International Development (AusAID), the Ford Foundation, IZA and the Chinese Foundation of Social Sciences.

1 Introduction

Social interactions, whether regular or occasional, influence individual decisions and behaviors. The effects of social networks on economic activity have been well documented (see Jackson, 2008; Ioannides, 2012; Jackson and Zenou, 2013; Jackson et al., 2014, for recent surveys), particularly in the labor market where social networks play an important role in transmitting information about jobs (Ioannides and Loury, 2004; Bayer et al., 2008; Topa, 2001, 2011). Social networks are also widely recognized to be very influential in migration decisions (Munshi, 2003; McKenzie and Rapoport, 2007, 2010; Beine et al. 2011a,b; Dolfin and Genicot, 2010; Bertoli and Fernández-Huertas Moraga, 2012). Nonetheless, little is known about the mechanisms by which networks exert such effects. The aim of this paper is to investigate the role of networks in great depth by disentangling the effect of strong and weak ties in migration decision.

There is a large body of literature on the role played by the different types of network in the labor market. In particular, Granovetter (1973, 1974, 1983) shows that weak ties are superior to strong ties in terms of providing support in getting a job.[1] Indeed, in a close network where everyone knows each other, information is shared and thus potential sources of information are quickly shaken down, whereby the network quickly becomes redundant in terms of access to new information. In contrast, Granovetter stresses the *strength of weak ties* involving a secondary ring of acquaintances who have contacts with networks outside the ego's network and therefore offer new sources of information about job opportunities.

In the present paper, we would like to investigate whether this is also true for migration decisions. Accordingly, we first derive a theoretical model to illustrate the different channels through which social networks may affect migration decisions. To be more precise, we consider a dynamic model in which individuals belong to different dyads and dyad members do not change over time. As a result, two individuals belonging to the same dyad hold a *strong tie* with each other. However, each dyad partner can meet other individuals outside the dyad partnership, referred to as *weak ties* or random encounters. By definition, weak ties are transitory and only last for one period.

Individuals can be in two different states: they have either migrated or not and, therefore, there will be three different types of dyads: both members have migrated, one member has migrated and the other has not migrated, or both members have not migrated. In this model, only workers who have migrated in the city can provide information about jobs to rural workers. Strong ties provide more reliable information than weak ties. Thus, information about jobs is essentially obtained through strong and weak ties and thus social networks.

[1] Granovetter (1973, 1974, 1983) defines *weak ties* in terms of lack of overlap in personal networks between any two agents, i.e. weak ties refer to a network of acquaintances who are less likely to be socially involved with one another. Formally, two agents A and B have a weak tie if there is little or no overlap between their respective personal networks. Vice versa, the tie is *strong* if most of A's contacts also appear in B's network.

We show that a unique steady-state equilibrium exists and explicitly determine the migration rate in the economy. We also show that the probability of migrating increases with the social interactions with strong ties, the social interactions with weak ties and the job-arrival rate, and decreases with the job-destruction rate in the city (or the emigration rate).

We subsequently test these theoretical results using a unique longitudinal dataset in China (RUMiC), where we observe the individuals and their networks before migration. As in the theoretical model, we define different types of networks based on the strength of the social interaction. First, we define strong ties based on the closest family contact and estimate the impact of the closet tie migrating on the individual's subsequent migration decisions. Second, we define weak ties as the share of previous migrants from the village. We measure the network at a point in time that precedes migration by between 1 and 3 years. There are several challenges when attempting to estimate the effects of networks on migration. First, endogeneity could arise because there could be unobservable factors affecting both the network characteristics and the migration decision. Selectivity could also arise to the extent that only individuals of a certain type self-select into certain networks. Second, there may be common shocks at the village levels that trigger migration. We address all these issues by providing different robustness checks.

Our results indicate that both weak and strong ties matter in the migration decision process, although the impact of weak ties is higher than that of strong ties. We also find that weak and strong ties act as complements in the migration decision. We show that if 50% of someone's weak ties have migrated, then his/her probability of migrating increases by 155% if his/her strong tie has migrated compared to the case in which the strong tie has not migrated. This indicates that the interactive effect between weak and strong ties is particularly strong above a certain threshold of the fraction of migrants in the village.

The structure of the paper is as follows. In the next section, we discuss the contribution of our paper compared to the literature on social networks and migration. In Section 3, we present the background context by describing the migration and the role played by networks in China. The theoretical model is developed in Section 4. The data and identification strategy are discussed in Section 5. In Section 6, we present our main empirical results. Several robustness checks are performed in Section 7. Finally, Section 8 concludes.

2 Related literature

There is a growing body of literature concerning social networks and migration, especially in developing countries. One important question concerns the extent to which the influence of networks is significant on top of the role of the traditional factors (such as the wage differential between the origin and the destination country, the bilateral distance between the two countries, etc.). The empirical literature based on structural gravity models (Beine et al., 2011a,b; Bertoli and Fernández-Huertas Moraga, 2012; Beine and Parsons 2012) finds an elasticity of about 0.4, which means that a ten percent increase in the bilateral migration stock will lead, on average, to a four percent increase in the bilateral migration flow over the next ten years.

At the microeconomic level, it is important to understand the exact role of networks in the migration decision. As noted by Dolfin and Genicot (2010), migrant networks can facilitate migration in three different ways: through providing information about the migration process itself; through providing information about jobs at the destination and aiding integration after arrival; and through helping to finance the costs of migration.

Recent work provides support for the role of networks in finding jobs at migrants' destinations. Using Mexican rainfall as an instrument for the size of migrants' US networks, Munshi (2003) finds that larger networks substantially improve Mexican immigrants' likelihood of US employment. The role of networks in alleviating migration costs has been investigated by McKenzie and Rapoport (2007, 2010), who find evidence suggesting that community networks tend to lower costs, especially for the less educated. Orrenius and Zavodny (2005) find that having a father or brother who has migrated to the US increases the likelihood of migration for males.[2]

In most of these papers, networks are measured by taking the share of migrants in the destination country from the same village of origin (see e.g. Munshi, 2003; McKenzie and Rapoport, 2007, 2010). However, this is clearly a very rough measure of social networks and one needs to open this black box to better understand the role of networks on migration. In the present paper, we measure weak ties in the same way but also include strong ties, defined as the closest contacts nominated by the head of the household. We show that one underestimates the effect of social networks on migration by not taking into account the strong ties in the mobility process. Moreover, we are able to show that there are strong complementarities between these two types of ties, especially above a certain threshold of the size of the weak ties. This is important because it means that both types of social interactions are key to understanding migration decision and that weak and strong ties reinforce each other.

Finally, we believe that our results shed light on the mechanisms by which networks encourage migration, because weak and strong ties provide different types of information in the migration process. Strong ties (measured here by the closest family contact) usually provide information about jobs at the destination and help (financial and other) to facilitate migration. Weak ties (as measured by the village in which the potential migrant resides) usually provide some information about jobs at the destination. Because weak and strong ties have different roles in encouraging migration, we believe that it is of paramount importance to disentangle between these two types of ties.

[2]See also Wahba and Zenou (2012).

3 Migration and social networks in China

3.1 Migration in China

Internal migration in China is important in magnitude and consequences. China is experiencing mass rural-urban migration, triggered by the economic reform that started at the end of the 1970s. Prior to that period, the combination of the household registration system (hukou) and the imposed quotas for per capita consumption considerably limited human mobility between rural and urban areas. Agricultural productivity increased with the beginning of the economic restructuring, yielding both an excess rural labor force and a more stable supply of food. Furthermore, these changes were accompanied by a rise in the inflow of foreign investment in urban areas, which itself created a high demand for low-priced labor force. The combination of these vicissitudes progressively generated the largest movement of labor in human history. Recent estimates reveal that over 160 migrant workers have moved from their rural residence to urban areas (NBS China, 2013).

While partially reformed, the hukou system persists and continues to influence the size and composition of the rural-to-urban migrant flows. Hukou regulations imply that migrants are only allowed to live in cities for few years and exclusively for working reasons. Furthermore, migrants often lack access to better paid jobs, social security or good schools for their children. In such a context in which most migrations are precarious in nature yet are also frequent and involve a large share of rural households, the role of social networks becomes crucial. For example, the network can provide information about job opportunities in the city, but also effective help (such as housing or financial support) in order to facilitate the move. In the next subsection, we provide a description about the main aspects of the social networks in China and how its role in influencing economic outcomes has been analyzed in the literature to date.

3.2 Social networks in China

Social connections – also known as *guanxi* – permeates many aspects of the Chinese culture and is frequently used to achieve the most disparate tasks in daily life, ranging from getting a job to providing favors. Given such pervasiveness, a growing number of economists have started to explore how the guanxi affects economic decisions and outcomes. A first set of studies look at the role of social networks in the context of the labor market. For example, Zhang and Zhao (2010) analyze the role of social networks in promoting the self-employment of migrants in Chinese cities. One of the features of their study is that they take into account the endogenous formation of migrants' networks after migration. Long et al (2013) examine the impact of networks on migrants' wages using the proportion of labor migrants in the home village as an indicator of the village social network.

The role of networks in the migration decision has previously been analyzed by Zhao (2003), who found that previous rural-to-urban migration – represented by the network of earlier migrants – positively influence subsequent migrations. However, the actual network is not observed in her study, and she relies on approximating it with the proportion of migrants from the same village who migrated in a given year. A similar network measure is used by Chen et al. (2009), who examine the role of networks in determining the cluster of migrants in certain destinations. Our paper contributes to the migration literature by studying the role played by different types of guanxi, namely weak and strong ties, where these networks are observed.

4 The theoretical model

4.1 Assumptions, notations and definitions

Consider a population of individuals of size one.

Dyads We assume that individuals belong to mutually exclusive two-person groups, referred to as *dyads*. We say that two individuals belonging to the same dyad hold a *strong tie* to each other. We assume that dyad members do not change over time. A strong tie is created once and for ever and can never be broken. Thus, we can consider strong ties as links between members of the same family, or between very close friends.

Individuals can be in either of two different states: have migrated (state 1) or have not migrated (state 0). Dyads, which consist of paired individuals, can thus be in three different states,[3] as follows:

(i) both members have migrated −we denote the number of such dyads by d_2;

(ii) one member has migrated and the other has not migrated (d_1);

(iii) both members have *not* migrated (d_0).

For example, a d_2 dyad means that both workers in the dyad (i.e. who have a strong tie relationship with each other) have migrated in the city, while a d_1 dyad means that one person in the dyad has migrated while the other has not.

[3]The inner ordering of dyad members does not matter.

Aggregate state By denoting the migration rate and the non-migration rate at time t by $m(t)$ and $n(t)$, where $m(t), n(t) \in [0,1]$, we have:

$$\begin{cases} m(t) = 2d_2(t) + d_1(t) \\ n(t) = 2d_0(t) + d_1(t) \end{cases} \tag{1}$$

The population normalization condition can then be written as

$$m(t) + n(t) = 1 \tag{2}$$

or, alternatively,

$$d_2(t) + d_1(t) + d_0(t) = \frac{1}{2} \tag{3}$$

Social interactions Time is continuous and individuals live forever. We assume repeated random pairwise meetings over time. Matching can take place between dyad partners or not. At time t, each individual can meet a weak tie with probability $\omega(t)$ and her strong-tie partner with probability $\alpha(t)$. We assume these probabilities to be constant and exogenous, not to vary over time and thus, they can be written as ω and α. We do not assume anything about ω and α being complement or substitute. We simply assume that individuals spend some time with their weak ties (captured by ω) and some time with their strong tie (captured by α).

We refer to matching inside the dyad partnership as *strong ties*, and to matching outside the dyad partnership as *weak ties* or random encounters. Information is exchanged within each matched pair as explained below.

Information transmission Workers migrate from the rural area to the urban area. We denote workers who have migrated (to the urban area) by workers of type m or $m-$workers or *migrants* and workers who have not migrated (and thus live in the rural area) by workers of type n or $n-$workers or *non-migrants*. In order to migrate, a rural worker needs to have some information about a job opportunity in the urban area. When the rural worker obtains this information about a job in the city, she may migrate to the city. Each piece of information about an opportunity is taken to arrive only to the $m-$workers, who can then direct it to one of their contacts who have not migrated (through either strong or weak ties). In other words, *only individuals who have already migrated and live in the city can help a rural individual to migrate to the city by providing her some information about an opportunity (about a job) in the urban area.* To be more precise, a worker of type m (a migrant) hears of a job opportunity in the city at the exogenous rate λ. This migrant will transmit this job opportunity to a non-migrant depending if the latter has a weak tie or a strong relationship with the migrant. Quite naturally, we assume that the quality of the job information is better when it comes from a strong tie than from a weak tie. To be more precise, a non-migrant will always migrate whenever she receives job information from her strong tie, while she will migrate with probability $0 < p < 1$ when the job information comes from a weak tie. This is because a strong tie, who has a long-term relationship with the non-migrant, always provides reliable information about jobs and can even provide financial help, or housing and help with other aspects of urban life for the migrant. On the other hand, a weak tie, who is a random encounter and has a short relationship with the non-migrant, can provide less reliable information about jobs.

We also assume that there is an exogenous rate δ that a $m-$worker goes back to the rural area. This is because migrants in the city can lose their jobs (at rate δ) and subsequently need to return to the rural area. This is particularly relevant in the case of China, since migrants without urban hukou are not allowed to stay in the city without working. Similarly, access to welfare programs is usually restricted to urban hukou holders. In other words, δ is the exogenous rate at which urban migrants return to the rural area. We call this the emigration rate or the (city) job-destruction rate. Therefore, migrants who hear about an opportunity in the city pass on this information to their current matched partner, who can be a strong or weak tie. Thus, information about opportunities in the city is essentially obtained through social networks.

This information transmission protocol defines a Markov process. The state variable is the relative size of each type of dyad. Transitions depend on labor market turnover and the nature of social interactions as captured by ω and α. Because of the continuous time Markov process, the probability of a two-state change is zero (small order) during a small interval of time t and $t+dt$. This means, in particular, that both members of a dyad cannot change their status at the same time. For example, two rural workers cannot migrate to the city at the same time, i.e. during t and $t+dt$, the probability assigned to a transition from a d_0-dyad to a d_2-dyad is zero. They can eventually both migrate, although it will take some time.

Flows of dyads between states It is readily checked that the net flow of dyads from each state between t and $t+dt$ is given by:

$$\begin{cases} \dot{d}_2(t) = (\alpha + \omega\, m(t)p)\, \lambda d_1(t) - 2\delta d_2(t) \\ \dot{d}_1(t) = 2\omega\, m(t)\lambda p d_0(t) - (\delta + \alpha\lambda + \omega\, m(t)\lambda p)\, d_1(t) + 2\delta d_2(t) \\ \dot{d}_0(t) = \delta d_1(t) - 2\omega\, m(t)\lambda p d_0(t) \end{cases} \tag{4}$$

Let us explain these equations in further detail, starting with the first one. Then, the variation of dyads composed of two $m-$workers ($\dot{d}_2(t)$) is equal to the number of d_1-dyads in which the $n-$worker has migrated to the city (through either her strong tie with probability

$\alpha\lambda$ or her weak tie with probability $\omega m(t)\lambda p$) minus the number of d_2-dyads in which one of the two migrants has returned to the rural area. In the second equation, the variation of dyads composed of one migrant and one non-migrant ($\dot{d}_1(t)$) is equal to the number of d_0-dyads in which one of the non-migrants has migrated to the city (only through her weak tie with probability $\omega\, m(t)\lambda p$ since her strong tie also lives in the rural area and therefore cannot transmit any opportunity in the city) minus the number of d_1-dyads in which either the migrant worker has returned to the rural area (with probability δ) or the rural worker has migrated to the city thanks to her strong or weak tie (with probability $[\alpha + \omega\, m(t)p]\,\lambda$) plus the number of d_2-dyads in which one of the two migrants has returned to the rural area. Finally, in the last equation, the variation of dyads composed of two rural workers ($\dot{d}_0(t)$) is equal to the number of d_1-dyads in which the migrant worker has returned to the rural area minus the number of d_0-dyads in which one of the rural workers has migrated to the city (only through her weak tie, with probability $\omega\, m(t)\lambda p$). These dynamic equations reflect the flows across dyads. Graphically,

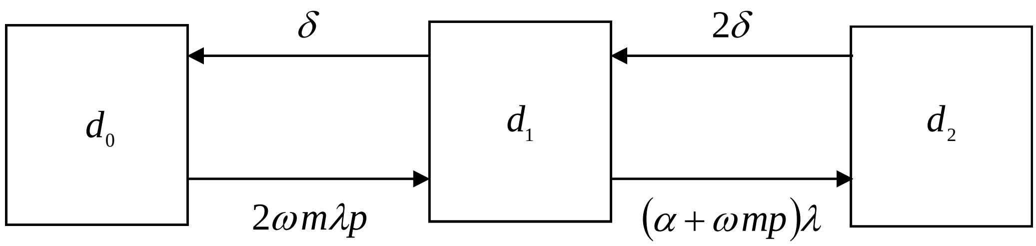

Figure 1: Flows between rural and urban areas.

4.2 Steady-state equilibrium

In a steady-state (d_2^*, d_1^*, d_0^*), each of the net flows in (4) is equal to zero. Setting these net flows equal to zero leads to the following relationships:

$$d_2^* = \frac{(\alpha + \omega m^* p)\lambda}{2\delta} d_1^* \tag{5}$$

$$d_1^* = \frac{2\omega m^* \lambda p}{\delta} d_0^* \tag{6}$$

where

$$d_0^* = \frac{1}{2} - d_2^* - d_1^* \tag{7}$$

$$m^* = 2d_2^* + d_1^* \tag{8}$$

$$n^* = 1 - m^* \tag{9}$$

Definition 1 *A steady-state labor market equilibrium is a four-tuple* $(d_2^*, d_1^*, d_0^*, m^*, n^*)$ *such that equations (5), (6), (7), (8) and (9) are satisfied.*

We have the following result:

Proposition 1

(i) *There always exists a steady-state equilibrium* $\mathcal{N}$ *where all rural individuals do not migrate to the city so that only* d_0-*dyads exist, that is* $d_2^* = d_1^* = m^* = 0$, $d_0^* = 1/2$ *and* $n^* = 1$.

(ii) *If*

$$\frac{\delta}{\lambda} < \frac{\omega p + \sqrt{\omega p \left(4\alpha + \omega p\right)}}{2} \tag{10}$$

there exists an interior steady-state equilibrium $\mathcal{I}$ *where* $0 < m^* < 1$ *is defined by*

$$m^* = \frac{\sqrt{\lambda \left(4\delta\alpha + \alpha^2\lambda + 2\alpha\omega\lambda p + \omega^2\lambda p^2\right)} - 2\delta - \alpha\lambda + \omega\lambda p}{2\omega\lambda p} \tag{11}$$

$0 < n^* < 1$ by (9), and $0 < d_0^* < 1/2$ is given by:

$$d_0^* = \frac{\delta^2}{\omega\lambda^2 p \left(\alpha + \omega p\right) + \omega\lambda p\sqrt{\lambda \left(4\delta\alpha + \alpha^2\lambda + 2\alpha\omega\lambda p + \omega^2\lambda p^2\right)}} \tag{12}$$

Furthermore, the other dyads are given by:

$$d_1^* = \frac{2m^*\omega\lambda p}{\delta} d_0^* \tag{13}$$

$$d_2^* = \frac{\omega\lambda^2 p \left(\alpha + \omega m^*\right) m^*}{\delta^2} d_0^* \tag{14}$$

If condition (10) holds, then an interior equilibrium always exists. Indeed, if the rate at which migrants return to the rural area δ is sufficiently low and/or the information-contact rate λ is sufficiently high, then an interior equilibrium exists. Otherwise, all workers will not migrate and stay in the rural area and the steady-state equilibrium $\mathcal{N}$ will prevail. Interestingly, in the interior steady-state equilibrium $\mathcal{I}$, we know m^*, the fraction of workers who have migrated to the city, and among them, those who have a strong tie migrating to the city (d_2^* dyads), as well as those who have a strong tie living in the rural area (d_1^* dyads).

4.3 Comparative statics

The key variable of our model is m^*, which is given by (11). Since we have a Markov process, m^* can have two equivalent interpretations: it is either the fraction of time each worker has spent migrating over their lifetime or *the unconditional probability of migrating for any worker in the steady state.* For the purpose of our empirical analysis, we use the second interpretation. As can be seen in (11), m^* is a function of α, the interaction with strong ties, ω, the interaction with weak ties, λ, the rate at which migrants hear about a job and δ, the job-destruction rate in the city (or the emigration rate). Let us now analyze the impact of these variables on the probability of migrating m^*. We have the following result:

Proposition 2 *Assume that condition (10) holds. Accordingly, the probability of migrating increases with the social interactions with strong ties α, the social interactions with weak ties and the job-arrival rate λ, and decreases with the job-destruction rate δ, that is*

$$\frac{\partial m^*}{\partial \alpha} > 0, \ \frac{\partial m^*}{\partial \omega} > 0, \ \frac{\partial m^*}{\partial \lambda} > 0, \ \frac{\partial m^*}{\partial \delta} < 0$$

Furthermore, the cross effect of weak and strong ties on the probability of migrating is undetermined, that is $\partial^2 m^* / \partial\alpha\partial\omega$ has an ambiguous sign.

This is an interesting result, which shows that both weak and strong ties can help a rural worker to migrate to the city. It also shows how the state of the economy in the city (captured by δ and λ) can affect the migration of workers.

4.4 Econometric equation

We would like to test Proposition 2, in other words, to evaluate the role of weak and strong ties in the migration decision of rural workers in China. Observe that m^* is defined by (11) as well as by $m^* = 2d_2^* + d_1^*$. Thus, using (5) and (6), we have:

$$m^* = \left[\frac{(\alpha m^*)\,\lambda\omega + (\omega m^*)^2\, p\lambda + \delta\,(\omega m^*)}{\delta} \right] \frac{2\omega\lambda p}{\delta} d_0^* \tag{15}$$

where d_0^* is defined by (12). In (15), we see that m^*, the probability of migrating is a function of ωm^*, the probability of migrating for weak ties and αm^*, the probability of migrating for strong ties, i.e. $m^* = f(\alpha m^*, \omega m^*, \lambda, \delta)$. This is what we want to test. Therefore, a (linear) reduced form of (15) can be written as:[4]

[4]In the empirical analysis, we will also test for a non-linear relationship between m_i and N_{is} and m_i and N_{iw}.

$$m_i = \beta_0 + \beta_1 N_{is} + \beta_2 N_{iw} + \beta_3 \lambda + \beta_4 \delta + \beta_5 N_{is} N_{iw} + \sum_j \beta_{ij} C_{ij} + \mathbf{X}'\theta + \varepsilon \quad (16)$$

where m_i is the (unconditional) probability of migrating for individual i, N_{is} is the fraction of i's strong ties who have migrated, N_{iw} is the fraction of i's weak ties who have migrated, λ measures the job-arrival rate in urban areas, δ measures the emigration rate, $k = S, W$ (S and W stand for strong and weak tie, respectively), C_{is} are the observable characteristics of the strong tie of individual i, while $\sum_j \beta_{ij} C_{ij}$ are the (average) characteristics of the workers belonging to the same village (weak ties) as individual i and $\mathbf{X}'$ are individual and household attributes of i, along with a regional indicator.

5 Data and identification

5.1 Data and descriptive evidence

Our analysis is based on the RUMiC data, collected as part of a large scale project conducted in China and comprising a rural household survey (RHS), an urban household survey and a migrant household survey. The migrants surveyed were randomly chosen from fifteen cities that are the top rural-urban migration destinations in China. Eight of these cities are in coastal regions (Shanghai, Nanjing, Wuxi, Hangzhou, Ningbo, Guangzhou, Shenzhen, and Dongguan), five are in central inland regions (Zhengzhou, Luoyang, Hefei, Bengbu, and Wuhan) and two are in the west (Chengdu and Chongqing). A sampling procedure was very carefully designed to ensure that migrants in the database constituted a representative sample of all the migrants in the fifteen cities.[5] For our purposes, we extract data from the 2008, 2009 and 2010 waves of the RHS (see Akgüc et al., 2014, for a technical description of the RUMiC panel dataset). The rural household survey (RHS) covers the principal migrant sending provinces and was conducted using the random samples from the annual China household income and expenditure surveys carried out in rural villages.[6] The survey contains detailed information about household members – including those currently migrating – and comprises socio-demographic characteristics, labor market outcomes, migration history and the family situation prior to leaving the hometown. The data also provide information about village characteristics, including the number of individuals who currently migrated, as well as other characteristics, such as total population, the number of self-employed individuals and the village average wage.

[5] See the RUMiCI Project's homepage (http://rumici.anu.edu.au/) for detailed documentation of the sampling method.

[6] The provinces covered by the RHS are: Anhui, Chongqing, Guangdong, Hebei, Henan, Hubei, Jiangsu, Sichuan, and Zhejiang.

In the data, we measure m_i by the probability that a given individual will migrate in one of the two years following the survey (conditioning on not having ever migrated). In other words, m_i is the probability of first-time migration, defined as an indicator that equals 1 if the individual will migrate in one of the two years following the 2008 survey, and conditioning on not having ever migrated before.

One of the important features of our analysis is that we model "future migration" as a function of "present covariates", and particularly network characteristics. The key explanatory covariates are based on two dimensions of the network: strength of social interactions (weak and strong ties) and migration status. We measure N_{is}, individual i's *strong ties who have migrated* by the fraction of the family's closest contacts who have migrated. Closest contacts are nominated by the head of the household or, in his/her absence, by the spouse. The head of the household could be individual i or someone closely related to him/her (the wife or children). In most of the empirical analysis, we will only consider the first contact nominated by the head of the household. In such a case, we will have a dummy variable that is equal to 1 if the strong tie has migrated and zero otherwise. We measure N_{iw}, individual i's *weak ties who have migrated*, by the fraction of individuals in the village where i resides who have migrated. In other words, weak ties represent the entire social space, which we measure at the village level.

We also have information concentring two additional parameters stemming from the theoretical model: the emigration rate (δ) and the rate at which job information reaches the network (λ). To capture these measures, we use two proxies. For δ, we use the village rate of labor immigration of individuals from outside the province (which is thought to relate to patterns of return migration), while for λ, we use the change in the urban unemployment rate in the capital city of the province in which the individuals live (which is arguably related to the rate at which information on job opportunities reaches the network).

Other covariates include additional characteristics of the networks. For example, for the village, we include the population size, the share of individuals who are self-employed and the average wage. For the closest contact of the family, we use indicators concerning the relationship status (being a relative vis-a-vis being a neighbor or a friend), the employment status (being a laborer or otherwise) and the education level (having ten or more years of education or otherwise).

Finally, we include a rich set of individual and household characteristics, as well as an indicator for whether the individual lives in a high emigration region.

For the purposes of our study, we restrict the sample to individuals aged 16 to 35. The reason is that more than two-thirds of first-time migrants move within this age window (the median age being 30 years) and we want to focus on individuals who move exclusively for labor-related reasons.[7]

[7]To corroborate our results, we perform several robustness checks, for example including individuals up

In Table 1, we present individual and household characteristics, as well as the network's attributes.

[*Insert Table* 1 *here*]

The table shows that the percentage of males is below 50%, reflecting that males of this age are more likely to have already migrated (and hence are excluded from our sample). About two-thirds of individuals are married, and have on average less than 9 years of education. Half of the individuals are the first child and are farmers. The average household size is above three, with less than one children aged 0-16 on average.[8] The table also reports income and wealth indicators of the household. Finally, nearly 60% of individuals live in provinces that send more migrants than they receive (which we define as emigration regions).

In terms of migration, we observe that 12.5% of those interviewed migrated the two years following the interview in 2008. This is our dependent variable m_i in (16).

In terms of network characteristics, on average, 17.4% of the weak ties migrated, while the proportion is slightly above 6% for strong ties. These large figures strongly reflect the magnitude of the migration phenomenon in rural China. The table also reports other characteristics of both weak and strong ties. For villages, the average population size is just over 2,500 individuals. The percentage of self-employed is rather low, perhaps reflecting that farming remains an important activity in these rural areas. The average monthly wage in the village is just above 400RMB. In terms of closest contact, we observe almost one-third are relatives, less than 20% have more than 10 years of education and nearly half of them work as laborers.

In Table 2, we provide some evidence about dyads where we only consider one strong tie, i.e. we take the first person nominated by the head of the household. In the theoretical model, we normalized the total population to 1. In the real world, this is not the case and we denote by T the total population of individuals in our sample who were surveyed in 2008 ($T = 2,192$). This does not include the strong ties. Since each person has one strong tie, the total population is $2T = 4,384$. Furthermore, in the theoretical model, we assume that $d_{10} = d_{01} = d_1$. This is not true in the data. Indeed, d_{10} is the number of dyads in which the person interviewed has migrated while her strong tie has not. Similarly, d_{01} is the number of dyads where the person interviewed has not migrated while her strong tie has. Note that each dyad has two persons so, for example, the number of persons interviewed who have migrated with a strong tie who has also migrated is d_2, while the total number of migrants who have migrant strong ties is $2d_2$. In Table 2, we see that $d_0 = 1,809$, which means that there are $1,809$ interviewed persons who have not migrated and whose strong ties have not

to 64 years and those who have already migrated before.

[8]The average household size in the total population 16-64 is also very similar (3.2), but the average number of children in the household is higher (1.9). This reflects the relatively lower average age of household members in our sample (27) vis-a-vis that of the total working age population (45).

migrated. We can also see that $d_{10} \neq d_{01}$, i.e. there are 247 interviewed persons who have migrated but their strong tie has not (d_{10}) while there are 110 interviewed individuals who have not migrated but their strong tie has (d_{01}). If we want to calculate the *unconditional* probability of migrating for an interviewed person, it is given by $(d_{10} + d_2)/T = 0.125$, which is the percentage given in Table 1. However, if we want to calculate m^*, as defined in the theoretical model by (11), we obtain:

$$m^* = d_{10}^* + d_{01}^* + 2d_2^* = \frac{d_{10}^*}{2T} + \frac{d_{01}^*}{2T} + \frac{2d_2^*}{2T} = 0.093$$

As a result, the *unconditional* probability of migrating for any person (including both interviewed individuals and their strong ties) is 9.3%, which is smaller than 12.5%, the *unconditional* probability of migrating for an interviewed individual. This is because there is less migration from strong ties with non-migrant interviewed individuals than from interviewed persons with non-migrant strong ties ($d_{10}^* = 247 > d_{10}^* = 110$).

[*Insert Table* 2 *here*]

Figures 2, 3 and 4 provide information about some characteristics of the strong ties. In Figure 2, we classify strong ties (or close ties) depending on the type of help they provide, distinguishing between financial help, psychological help, help with daily affairs and no help. This figure shows that close ties who have migrated are more likely to provide financial help than those who have not migrated, while the reverse is true for psychological help. This indicates that the migrant strong ties more prone to supplying material and financial support (and not just information) to the potential migrant than those who have not migrated.

[*Insert Figure* 2 *here*]

In Figure 3, we depict the frequency of close tie contacts. There are stark differences depending on the migration status of the strong tie. Close ties who have migrated have somewhat less frequent contact with individuals in rural areas than those who have migrated. This is certainly due to the fact that the geographical distance between strong ties is less for those who have not migrated than for the migrants.

[*Insert Figure* 3 *here*]

Finally, we report some statistics about money and gifts exchange in Figure 4. While, the amount of money received is generally larger for both close ties who did not migrate and those who did, the difference is much more marked for the latter group. This complements the information about financial help reported in Figure 2.

[*Insert Figure* 4 *here*]

5.2 Econometric strategy and identification

There are several challenges when trying to estimate network effects, mostly related with the endogeneity of the network.

First, *network endogeneity of strong ties* might arise because there could be unobservable factors affecting both friendship formation and migration decision.[9] Moreover, since individuals only name up to five persons as their strong ties and we do not observe the whole network (i.e. strong ties have not been interviewed), the problem of selectivity into networks is similar to the endogeneity of choosing one's closest tie. To deal with this issue, we perform the three following robustness checks. First, we narrow down our sample to only strong ties who are known to the interviewed individual since age 16. In this case, it is hard to believe that the reason why two 16 years old become friends is because they anticipate that they will help each other to migrate 10 years after (the average age of migration in our sample is 27). Second, remember that the question about strong ties is only asked to the head of household. As a robustness check, we only consider individuals in the family who are *not* the head of the household since, in that case, strong ties are all distant 2 from the individual who considers migrating, i.e. they are the closest contacts of their relatives (i.e. the head of the household). For example, if we consider the son of the head of the household (the father), then we only look at the closest contacts of the father, who are distant 2 from the son. In that case, the choice of friendship is not entirely endogenous since it is chosen by the father but still indirectly affects the migration decision of the son. Finally, we narrow down the sample to only strong ties who know the interviewed individual since age 16 and individuals who are not the head of the household. As a consequence, it is difficult to argue that the head of the household's friendship decision made at the age of 16 can affect the migration decision of the children, since the latter were not even born!

Second, there may be common shocks at the village levels that trigger migration (see e.g. Munshi, 2003). Indeed, there may be a drought in the village that causes me and my closest contact to migrate and this has nothing to do with peer effects. To deal with this issue, we include the characteristics of the village to which each individual belongs.

[9] Observe that the reflection problem (Manski, 1993) does not arise here because the reference group is different from one individual to the other. Indeed, the best friend(s) nominated by the head of household is (are) usually *not* the same across different households, especially when people nominate at least three closest contacts. This is a similar approach to the network literature where the presence of intransitive triads in the network solves the reflection problem (Lee, 2007; Bramoullé et al., 2009; Calvó-Armengol et al., 2009; Blume et al., 2011).

6 Benchmark results

The aim of our empirical analysis is to examine the effects of both weak and strong ties in the individual migration decisions, as well as their potential interactive effects. Accordingly, we estimate equation (16).

The results in Table 3 present the benchmark case. All estimates represent marginal effects of a probit model. The estimates of the individual and household covariates produce expected results. For example, being a male and young individual have a positive impact on migration. Other positive factors are the household size and the farmer status and land size, as well as residing in a high emigration region. On the other hand, higher income and a higher value of the house are negatively correlated with the decision of migration. The estimates of the parameters δ and λ yield results that match the theoretical model's predictions. Higher return migration has a detrimental effect on first-time migration (although the effect is only statistically significant in a few models). Similarly, increasing unemployment reduces the rate at which job information reaches the network, thereby reducing migration.

Let us now focus on the role of weak and strong ties in migration.

Besides individual and household covariates Column I also includes the migration status of the weak ties, which is the village share of migrants who have migrated, i.e. N_w. This is the standard measure of social network in the migration literature (see Munshi, 2003; McKenzie and Rapoport, 2007, 2010; Dolfin and Genicot, 2010). We find that the estimate is positive and highly statistically significant. The magnitude of the estimate implies that a one percent increase in N_w is associated with an increase in the probability of migration by about 0.18 percentage points.

In column II, we estimate a model that only includes the impact of a strong-tie's migration status (N_s) on the decision to migrate. While the estimate is only significant at the 10% level, the magnitudes is rather sizeable. Having a close friend who migrated is associated with an increased probability of migration by about 5 percentage points.

In the third column, we show estimates from a model in which we include both type of networks (weak and strong ties). Estimates are very similar and are, in fact, not statistically different from those in columns I and II. In column IV, we control for additional characteristics of the network. Estimates of N_w are slightly lower than those in the previous models, while estimates of N_s are somewhat higher (and estimated with higher statistical precision).[10] Overall, the pattern is very similar and both strong and weak ties are strongly and positively related with the future migration decision. In columns I to IV, we estimate the econometric equation (16) with $\beta_5 = 0$.

[10]While village population and self-employment seem not to matter substantially in the migration decision, the village wage acts as a retention factor, i.e. the higher the wage is, the lower are the incentives to leave the village. In terms of strong ties, the more educated the tie is, the less likely is the probability of the individual to leave the village.

[*Insert Table* 3 *here*]

In column V, we interact the strong and weak tie variables and thus include $\beta_5 N_{is} N_{iw}$ in the econometric equation, as in (16). The estimates reveal a very interesting pattern: the effect of N_s disappears (with an estimate that is both statistically and economically negligible). However, this does not mean that the total effect of strong ties on migration is insignificant. Indeed, from equation (16), we can see that:

$$\frac{\partial m_i}{\partial N_{is}} = \beta_1 + \beta_5 N_{iw} = -0.0101 + 0.2845 N_{iw}$$

From Table 1, the average value of N_{iw} is 0.174, so that the net effect is

$$\frac{\partial m_i}{\partial N_{is}} = -0.0101 + 0.2845 \times 0.174 = 0.0394 \qquad (17)$$

which is roughly what we obtain in column III. The effect of weak ties on migration is relatively similar to those previously obtained. The new result in column 5 is the fact that $\beta_5 = \partial^2 m_i / \partial N_{iw} \partial N_{is}$ is strictly positive and statistically significant. This means that *weak and strong ties act as complements in the migration decision*. In other words, the higher the fraction of villagers who migrate, the greater the effect of migrants' strong ties on the individuals' migration decision. Column VI tests for non-linear effects of weak ties. Interestingly, we find that the effect of weak ties on migration increases up to a certain threshold and subsequently decreases. This result resembles the idea of network congestion observed in the case of labor networks in Egypt by Wahba and Zenou (2005). However, the addition of the quadratic term leaves the estimates of the interaction effect virtually unchanged.

If we compare column I, which is the standard way in which researchers have tested the role of networks on migration decision (Munshi, 2003; McKenzie and Rapoport, 2007, 2010) and column VI, where we have the role of both weak and strong ties as well as their interactions, we see that the coefficient of the weak ties is divided by almost two (from 0.18 to 0.33). Indeed, if we proceed as in (17), albeit for weak ties, using column VI and Table 1, we obtain:

$$\frac{\partial m_i}{\partial N_{iw}} = \beta_2 + \beta_5 N_{is} = 0.3128 + 0.2761 \times 0.062 = 0.33$$

This means that by only regressing the fraction of villagers that migrate (weak ties) on the probability of migrating, we *underestimate* the real effect of social networks on migration decision.

In order to quantify the impact of weak versus strong ties on the migration decision, we report in Table 4 the predicted probabilities of migration based on the estimations in Table

3. Let us start with the estimations of column IV where the interaction term is not included ($\beta_5 = 0$). First, we observe that when nobody has migrated in the village ($N_w = 0$), then the effect of strong tie on migration is important. Indeed, 14.77% of individuals migrate when their strong tie has also migrated, while only 8.65% will migrate if their strong tie did not migrate. More importantly, if we compare weak and strong ties, one needs to reside in a village where 35% of people have migrated to roughly obtain a migration rate of 15% when the strong tie has not migrated. In other words, the impact of a strong tie who has migrated on a person's own migration decision is "equivalent" to the impact of 35% of the people from the village who have migrated. More generally, if we double the percentage of weak ties that have migrated from 30 to 60%, when the strong tie has not migrated, it increases the person's own probability of migrating by 50% (i.e. from 14% to 21%).

[*Insert Table* 4 *here*]

Figure 5 shows these results for all possible values of weak-tie migrants. We see that the increase difference in migration probability between a strong tie who has migrated and not is relatively constant when the share of weak ties that migrate increases. This is clearly not the case when there are interaction effects.

[*Insert Figure* 5 *here*]

Let us now look at the second part of Table 4 based on estimations of column V of Table 3, where the interaction term is introduced. If we consider Figure 6, we see strong complementarities between weak and strong ties. If 10% of my weak ties have migrated, then my probability of migrating does not significantly increase (from 10.48 to 12.55%) whether or not my strong tie has migrated. On the contrary, if 50% of my weak ties have migrated, then my probability of migrating increases by 155% (from 17% to 43.37%) if my strong tie has migrated compared to the case when he/she has not migrated. This means that weak and strong ties are complements above a certain threshold. In other words, the interactive effect between weak and strong ties is particularly strong for values of N_w in the third and fourth quartile.

[*Insert Figure* 6 *here*]

7 Robustness checks

In this section, we perform several robustness checks. In section 5.2, we indicate that the main threat to identification is the *network endogeneity of strong ties*, since individual may form friendship relationships that help them to migrate. To address this issue, we first only consider individuals in the family who are not the head of the household because, in such a case, strong ties are distant 2 from them, i.e. they are closest contacts of their relatives (i.e. the head of the household). For example, if we consider the son in the family who considers migrating and the head of the household is the father, then we will look at the impact of the migration of the closest contact of the father on the son's decision to migrate. Table 5 displays the results. By excluding this small sample of individuals (less than 200), the results do not substantially change the pattern of estimates, with a strong effect of both weak and strong ties in the model without interaction, as well as a consistent evidence of positive complementarity in the model in column V. As in the previous section, we can calculate the effect of strong ties on migration when the interaction terms is introduced (column V). Proceeding as in (17), we obtain:

$$\frac{\partial m_i}{\partial N_{is}} = 0.041$$

which is very close to the benchmark model (where the value was 0.039).

[*Insert Table* 5 *here*]

As a second robustness check, we narrow down our sample to only strong ties who have known the interviewed individual since the age of 16. Once again, this robustness check aims to exclude the potential endogeneity in the choice of the strong tie. As can be seen from Table 6, the pattern of the results is unchanged. As above, proceeding as in (17), we obtain:

$$\frac{\partial m_i}{\partial N_{is}} = 0.057$$

which is again close to the benchmark value.

[*Insert Table* 6 *here*]

To address the endogeneity of friendship decision on migration decision in an even more convincing manner, we consider both individuals in the family who are not the head of the household and friendship relationships (strong ties) that date back to when the head of the household was 16. For example, if we consider the son in the family who considers migrating and the head of the household is the father, then we will look at the impact of the migration of the closest contact of the father on the son's decision to migrate, given that this closest contact already knew the father when the latter was 16 and (very likely) the son was not even born. Table 7 displays the results, from which we can see that all the main results remain unchanged.

[*Insert Table* 7 *here*]

So far, we have restricted the sample of workers to those aged between 16 and 35. As a robustness check, in Table 8, we include all individuals aged 16 to 64. Not surprisingly, the magnitude of the effects of weak and strong ties are "diluted" once additional individuals (who are less likely to migrate and less likely to be in a network of migrants) are included, although the results are globally robust.

[*Insert Table* 8 *here*]

In Table 9, we extend the sample of individuals aged 16-35 to those who have previous migration experience. Once again, not surprisingly, the network effect is much stronger compared to the benchmark model. One possibility for this large effect is that migration networks are highly endogenous. Indeed, thanks to previous migration, the individual has contributed to the network status of his/her network, and hence a correlation between today's migration and network status is potentially affected by reverse causality, leading to upwardly biased estimates. This final check confirms the importance of restricting our analysis to first-time migrants.

[*Insert Table* 9 *here*]

We also performed a series of additional tests. In particular, we tested for the definition of strong ties using information on *the closest three ties* instead of only the first strong tie. In this case, N_s is just the fraction of strong ties who have migrated. Table 10 reports the results[11] which are similar to those of our benchmark analysis (Table 3), despite being statistically weaker. In Table 10, we also explored a model that introduces an additional non-linear effect in the strong tie (which is possible now since individuals have more than one strong tie), that is N_s^2. The results are not affected and we see that the effect of strong tie on migration decision is increasing and convex.

[*Insert Table* 10 *here*]

In Table 11, we introduce interaction terms between all characteristics of strong ties (C_s) and the weak ties. The addition of these interactions leaves the main estimates unchanged.

[*Insert Table* 11 *here*]

Finally, in Table 12, we estimate models in which we change the definition of our dependent variable. Indeed, we explore alternative definitions of our endogenous variable, in which migration is defined from rural to urban areas within the local province or to another province. The first alternative definition includes migrations from rural to rural or rural to urban areas, and hence reflects the broadest definition. The second definition is migration from rural to urban area, albeit within the local county. The third definition is migration from rural to urban area in another province, and hence is the strictest definition. Note that the change of migration definition applies to both our outcome variable and the strong tie

[11] For the ease of the exposition, we do not report the results for the control variables, which are the same as in the previous tables.

variable. Interestingly, our results (displayed in Table 12) hold and are robust to the change in the definition of migration.

[*Insert Table* 12 *here*]

8 Conclusion

It is well established that migration experience among an individual's family and community networks tends to encourage migration. However, there exists little research investigating the mechanisms by which networks exert such effects.

We study these mechanisms by first developing a theoretical model that illustrates the different channels through which networks may affect migration decisions. Using unique data from China, we subsequently estimate the role played by social networks in the migration decision using networks prior to migration, distinguishing between strong and weak ties. Strong ties are measured by the closest contact nominated by a given individual, while weak ties are determined by the fraction of migrants from the village in which the individual resides.

Our results indicate that both weak and strong ties matter in the migration decision process, although the impact of weak ties is higher than that of strong ties. We also show that one underestimates the effect of social networks on migration by not taking into account the strong ties in the mobility process. We finally find that weak and strong ties act as complements in the migration decision, which indicates that the interactive effect between weak and strong ties is particularly strong above a certain threshold of the size of weak ties.

We believe that our results shed light on the mechanisms by which networks encourage migration, because weak and strong ties provide different types of information in the migration process. Strong ties (as measured by the closest family contact as here) usually provide information on jobs at the destination and different types of help to migrate or upon migration. In particular, as seen in Figure 2, migrant strong ties do provide much more financial help than non-migrant strong ties. Weak ties (as measured by the village in which the potential migrant resides) usually provide some information about jobs at the destination. Given that weak and strong ties play different roles in encouraging migration and because they complement each other, we believe that it is of paramount importance to disentangle between these two types of ties. Thus, our paper provides a first attempt at answering a very important question in relation to both social networks and migration, but clearly more research is needed to better understand the relationship between the two.

References

[1] Akgüc, M., Giulietti C. and K.F. Zimmermann (2014), "The RUMiC longitudinal survey: Fostering research on labor markets in China," *IZA Journal of Labor and Development*, forthcoming.

[2] Bayer, P., Ross, S.L. and G. Topa (2008), "Place of work and place of residence: Informal hiring networks and labor market outcomes," *Journal of Political Economy* 116, 1150-1196.

[3] Beine, M , Docquier, F. and C. Özden (2011a), "Diasporas," *Journal of Development Economics* 95, 30-41.

[4] Beine, M , Docquier, F. and C. Özden (2011b), "Dissecting network externalities in international migration," CESifo Working Paper No. 3333.

[5] Beine, M. and C. Parsons (2012), "Climatic factors as determinants of international migration," CESifo Working Paper No. 3747.

[6] Bertoli, S. and J. Fernández-Huertas Moraga (2012), "Visa policies, networks and the cliff at the border," IZA Discussion Paper No. 7094.

[7] Blume, L.E., Brock, W.A., Durlauf, S.N. and Y.M. Ioannides (2011), "Identification of social interactions," In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Publisher, pp. 853-964.

[8] Bramoullé, Y., Djebbari, H. and B. Fortin (2009), "Identification of peer effects through social networks," *Journal of Econometrics* 150, 41-55.

[9] Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009), "Peer effects and social networks in education," *Review of Economic Studies* 76, 1239-1267.

[10] Chen Y., Jin G.Z, and Yue Y. (2010), "Peer migration in China, " NBER Discussion Paper No. 15671.

[11] Dolfin, S. and G. Genicot (2010), What do networks do? The role of networks on migration and "coyote" use," *Review of Development Economics* 14, 343-359.

[12] Granovetter, M.S. (1973), "The strength of weak ties," *American Journal of Sociology* 78, 1360-1380.

[13] Granovetter, M.S. (1974), *Getting a Job: A Study of Contacts and Careers*, Cambridge, MA: Harvard University Press.

[14] Granovetter, M.S. (1983), "The strength of weak ties: A network theory revisited," *Sociological Theory* 1, 201-233.

[15] Ioannides, Y.M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.

[16] Ioannides, Y.M. and D.L. Loury (2004), "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature* 42, 1056-1093.

[17] Jackson, M.O. (2008), *Social and Economic Networks*, Princeton: Princeton University Press.

[18] Jackson, M.O., Rogers, B.W. and Y. Zenou (2014), "The impact of social networks on economic behavior," Unpublished manuscript, Stanford University.

[19] Jackson, M.O. and Y. Zenou (2013), *Economic Analyses of Social Networks*, The International Library of Critical Writings in Economics, London: Edward Elgar Publishing.

[20] Lee, L.F. (2007), "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics* 140, 333-374.

[21] Long, W., Appleton, S. and L. Song (2013), "Job contact networks and wages of rural-urban migrants in China," IZA Discussion Paper No. 7577.

[22] Manski, C.F. (1993), "Identification of endogenous effects: The reflection problem," *Review of Economic Studies* 60, 531-542.

[23] McKenzie, D. and H. Rapoport (2007), "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics* 84, 1-24.

[24] McKenzie, D. and H. Rapoport (2010), "Self-selection patterns in Mexico–US migration: The role of migration networks," *Review of Economics and Statistics* 92, 811-821.

[25] Munshi, K. (2003), "Identification of network effects: Mexican migrants in the US Labor market," *Quarterly Journal of Economics* 118, 549-597.

[26] NBS China (2013), "China Statistical Yearbook 2010, " National Bureau of Statistics of China, China Statistics Press.

[27] Orrenius, P. and M, Zavodny (2005), "Self-selection among undocumented immigrants from Mexico," *Journal of Development Economics* 78, 215-240.

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

[29] Topa, G. (2011), "Labor markets and referrals," In: J. Benhabib, A. Bisin and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Science, pp. 1193-1221.

[30] Wahba, J. and Y. Zenou (2005), "Density, social networks and job search methods: Theory and application to Egypt," *Journal of Development Economics* 78, 443-473.

[31] Wahba, J. and Y. Zenou (2012), "Out of sight, out of mind: Migration, entrepreneurship and social capital," *Regional Science and Urban Economics* 42 890-903.

[32] Zhang, J. and Z. Zhao (2011), "Social-family network and self-employment: Evidence from temporary rural-urban migrants in China," IZA Discussion Paper No. 5446.

[33] Zhao, Y. (2003), "The role of migrant networks in labor migration: The case of China," *Contemporary Economic Policy* 21, 500-511.

Appendix

Proof of Proposition 1: We establish the proof in two steps. First, Lemma 1 characterizes all steady-state dyad flows. Lemma 2 then provides conditions for their existence.

Lemma 1 *There exists at most two different steady-state equilibria:* (i) *a non-migration equilibrium* $\mathcal{N}$ *such that* $m^* = 0$ *and* $n^* = 1$*,* (ii) *an interior equilibrium* $\mathcal{I}$ *such that* $0 < m^* < 1$ *and* $0 < n^* < 1$*.*

Proof: By combining (5) to (8), we easily obtain:

$$m^* = [(\alpha + \omega m^* p)\lambda + \delta] \frac{2\omega m^* \lambda p}{\delta^2} d_0^* \tag{18}$$

We consider two different cases.

(i) If $m^* = 0$, then equation (18) is satisfied. Furthermore, using (5) and (6), this implies that $d_1^* = d_2^* = 0$ and, using (7) and (9), we have $d_0^* = 1/2$ and $n^* = 1$. This is referred to as steady-state $\mathcal{N}$.

(ii) If $m^* > 0$, then solving equation (18) yields:

$$m^* = \frac{1}{\lambda \omega p} \left[\frac{\delta^2}{2\omega\lambda p d_0^*} - \delta \right] - \frac{\alpha}{\omega p}$$

Define $Z = \alpha / (\omega p)$, $B = \delta / (\omega \lambda p)$. This equation can now be written as:

$$m^* = \frac{B^2}{2d_0^*} - B - Z > 0 \tag{19}$$

Moreover, by using (5) and (6), we obtain:

$$d_1^* = \frac{2m^*}{B} d_0^* \ , \qquad d_2^* = \frac{(Z + m^*)\, m^*}{B^2} d_0^* \tag{20}$$

• Let us first focus on the case where $m^* = 1$. In that case, it has to be that only d_2-dyads exist and thus $d_0^* = d_1^* = 0$, which, using (20) implies that: $d_2^* = 0$. So this case is not possible.

• Let us now thus focus on the case: $0 < m^* < 1$ (which implies that $0 < n^* < 1$).

By plugging (19) and (20) into (7) and after some algebra, we obtain that d_0^* solves $\Phi(d_0^*) = 0$ where $\Phi(d_0^*)$ is the following second-order polynomial:

$$\Phi(d_0^*) = -\frac{Z}{B} d_0^{*2} - \frac{(1+Z)}{2} d_0^* + \left(\frac{B}{2}\right)^2 = 0 \tag{21}$$

This completes the proof of the lemma. ▮

Lemma 2

(i) The steady-state equilibrium $\mathcal{N}$ always exists.

(iv) The steady-state equilibrium $\mathcal{I}$ exists when

$$\frac{\delta}{\lambda} < \frac{\omega p + \sqrt{\omega p \left(4\alpha + \omega p\right)}}{2}$$

Proof.

(i) In this equilibrium $m^* = 0$. There are only d_0-dyads. So when a d_0-dyad is formed it is never destroyed and thus this equilibrium is always sustainable.

(ii) We know from Lemma 1 that a steady-state $\mathcal{I}$ exists and that $m^* \neq 1$. We now have to check that $m^* > 0$ and $0 < d_0^* < 1/2$. Let us thus verify whether there exists some $0 < d_0^* < 1/2$ such that $\Phi(d_0^*) = 0$, where $\Phi(\cdot)$ is given by (21). We have $\Phi(0) = \left(B/2\right)^2 > 0$ and $\Phi'(0) = -\left(1+Z\right)/2 < 0$. Therefore, (21) has a unique positive root smaller than $1/2$ if and only if

$$\Phi(1/2) = \frac{1}{4}\left[B^2 - (1+Z) - \frac{Z}{B}\right] = \frac{1}{4}(1+\frac{1}{B})(B^2 - B - Z) < 0.$$

The unique positive solution to $B^2 - B - Z = 0$ is $\left[1+\sqrt{1+4Z}\right]/2$, which, using the value of Z, is equal to: $\left[1+\sqrt{1+4\alpha/\left(\omega p\right)}\right]/2$. Then, $d_0^* < 1/2$ if and only if $B < \left[1+\sqrt{1+4\alpha/\left(\omega p\right)}\right]/2$, equivalent to:

$$\frac{\delta}{\lambda} < \frac{\omega p + \sqrt{\omega p \left(4\alpha + \omega p\right)}}{2}$$

Observe that $d_0^* < 1/2$ guarantees that $m^* > 0$.

Finally, to obtain (11) and (12), we proceed as follows. First, we plug the values of d_2^* and d_1^* from (5) and (6) into $m^* = 2d_2^* + d_1^*$ to obtain (18) or equivalently:

$$m^* = \left[\frac{(\alpha + \omega m^* p)\lambda}{\delta} + 1\right]\frac{2\omega m^* \lambda p}{\delta} d_0^* \tag{22}$$

Then, we plug the values of d_2^* and d_1^* from (5) and (6) into $d_0^* = \frac{1}{2} - d_2^* - d_1^*$ to obtain:

$$d_0^* = \frac{\delta^2}{2\delta^2 + \left[(\alpha + \omega m^* p)\lambda + 2\delta\right]\omega m^* p \lambda} \tag{23}$$

By solving simultaneously equations (22) and (23), we get (11) and (12). ▮

Proof of Proposition 2: By differentiating (11), we obtain:

$$2\omega\lambda p\frac{\partial m^*}{\partial\alpha} = \frac{2\delta\lambda+\alpha\lambda^2+\omega\lambda^2 p}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}} - \lambda \tag{24}$$

Thus

$$\frac{\partial m^*}{\partial\alpha} > 0 \Leftrightarrow 2\delta+\alpha\lambda+\omega\lambda p > \sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}$$

It is easily verified that this inequality is equivalent to $\delta+\omega\lambda p>0$, which is always true.

By differentiating (11), we obtain:

$$\begin{aligned} 2\lambda p\omega^2\frac{\partial m^*}{\partial\omega} &= \left[\frac{\lambda^2 p\left(\alpha+\omega p\right)}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}+\lambda p\right]\omega \\ &\quad -\left[\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}-2\delta-\alpha\lambda+\omega\lambda p\right] \end{aligned}$$

Thus

$$\begin{aligned} \frac{\partial m^*}{\partial\omega} &> 0 \Leftrightarrow \left[\frac{\lambda^2 p\left(\alpha+\omega p\right)}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}+\lambda p\right]\omega \\ &> \sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}-2\delta-\alpha\lambda+\omega\lambda p \end{aligned}$$

which is equivalent to:

$$\left(2\delta+\alpha\lambda\right)\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)} > 4\delta\alpha\lambda+\alpha^2\lambda^2+\alpha\omega\lambda^2 p$$

That is

$$\left(2\delta+\alpha\lambda\right)^2\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right) > \left(4\delta\alpha\lambda+\alpha^2\lambda^2+\alpha\omega\lambda^2 p\right)^2$$

It is easily verified that this inequality is always true.

Let us now calculate $\frac{\partial^2 m^*}{\partial\alpha\partial\omega}$. By differentiating (24), we obtain:

$$2\lambda p\frac{\partial^2 m^*}{\partial\alpha\partial\omega} =$$

$$= \frac{\omega\lambda^2 p\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)} - \frac{\left(2\delta\lambda+\alpha\lambda^2+\omega\lambda^2 p\right)\left[\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)+\omega\lambda^2 p(\alpha+\omega p)\right]}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}}{\omega^2\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)} + \frac{\lambda}{\omega^2}$$

This implies that

$$\frac{\partial^2 m^*}{\partial\alpha\partial\omega} \gtreqless 0$$

$$\Leftrightarrow \quad \frac{\omega\lambda^2 p\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}-\frac{\left(2\delta\lambda+\alpha\lambda^2+\omega\lambda^2 p\right)\left[\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)+\omega\lambda^2 p(\alpha+\omega p)\right]}{\sqrt{\lambda(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2)}}}{\omega^2\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}$$

$$+\frac{\lambda}{\omega^2} \quad \gtreqless \quad 0$$

This is equivalent to:

$$\frac{\omega\lambda p\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}+\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}{\omega^2\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}$$

$$\gtreqless \quad \frac{\left(2\delta\lambda+\alpha\lambda^2+\omega\lambda^2 p\right)\left[\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)+\omega\lambda^2 p\left(\alpha+\omega p\right)\right]}{\omega^2\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}$$

or equivalently:

$$\begin{aligned}
&\omega\lambda^2 p\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)\\
&+\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}\\
\gtreqless\ &\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)\left(2\delta\lambda+\alpha\lambda^2+\omega\lambda^2 p\right)\\
&+\omega\lambda p\left(2\delta\lambda+\alpha\lambda^2+\omega\lambda^2 p\right)\left(\alpha+\omega p\right)
\end{aligned}$$

It is clearly impossible to sign this inequality and thus $\frac{\partial^2 m^*}{\partial\alpha\partial\omega}$ has an ambiguous sign.

By differentiating (11), we obtain

$$\begin{aligned}
2\omega p\lambda^2\frac{\partial m^*}{\partial\lambda} \quad = \quad &\frac{2\delta\alpha\lambda+\alpha^2\lambda^2+2\alpha\omega\lambda^2 p+\omega^2\lambda^2 p^2}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}\\
&-\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}+2\delta
\end{aligned}$$

Thus

$$\frac{\partial m^*}{\partial\lambda}>0$$

$$\Leftrightarrow \frac{2\delta\alpha\lambda+\alpha^2\lambda^2+2\alpha\omega\lambda^2 p+\omega^2\lambda^2 p^2}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}+2\delta>\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}$$

This is equivalent to

$$\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}>\alpha\lambda$$

which are clearly always true .

Finally, by differentiating (11), we obtain

$$2\omega\lambda p\frac{\partial m^*}{\partial\delta}=\frac{2\alpha\lambda}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}-2\delta$$

Thus

$$\frac{\partial m^*}{\partial\delta}<0\Leftrightarrow\frac{\alpha\lambda}{\sqrt{\lambda\left(4\delta\alpha+\alpha^2\lambda+2\alpha\omega\lambda p+\omega^2\lambda p^2\right)}}<\delta$$

It is easily verified that this inequality is always true. ∎

Tables and Figures

Table 1: Summary statistics

Personal and household characteristics		**Migration**	
Male (D)	0.435	Will migrate in the following two years	0.125
	(0.496)		(0.330)
Age	27.137		
	(5.536)	**Network characteristics**	
Marital status (D)	0.623	N_w: Village share of migrants	0.174
	(0.485)		(0.132)
Years of education	8.676	N_s: Close tie migrated	0.062
	(2.614)		(0.241)
Eldest child (D)	0.454	C_w: Village population	2.555
	(0.498)		(1.999)
Farmer worker (D)	0.520	C_w: Village self-employed	0.017
	(0.500)		(0.022)
Household size	3.462	C_w: Village average wage	410.185
	(1.127)		(255.448)
Household's number of children	0.775	C_s: Close tie relative	0.362
	(0.831)		(0.481)
Housing value (log RMB)	10.011	C_s: Close high education	0.146
	(0.607)		(0.354)
Size of family farm land (Mu)	10.366	C_s: Close tie laborer	0.455
	(1.070)		(0.498)
Household income (log RMB)	1.334	δ: Rate of village immigration	0.062
	(0.781)		(0.179)
Emigration region (D)	0.566	λ: Change in urban unemployment rate	(0.139)
	(0.496)		(0.101)

Observations: 2192. Source: RUMiC data, RHS waves 2008, 2009 and 2010. Standard deviations errors in parenthesis.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak tie who migrated.
C_s / C_w refer to other characteristics of the strong / weak tie.
Strong tie refers to closest person in network; Weak tie refers to village.

Table 2: Dyads

		Close tie		
		Not migrated	Migrated	Total
Individual	Does not migrate	1,809	110	1,919
	Migrates	247	26	273
	Total	2,056	136	2,192
	d_0:	=1809/2192	0.825	
	d_{10}:	=247/2192	0.113	
	d_{01}:	=110/2192	0.050	
	d_2:	=136/2192	0.012	

Source: RUMiC data, RHS waves 2008, 2009 and 2010.
Entries in the table refers to cross-tabulation of the main dependent variable and the migration status of the closest tie.

Table 3: Benchmark results

	I	II	III	IV	V	VI
N_w: Village outmigrants	0.1829***		0.1792***	0.1545***	0.1312***	0.3128***
	(0.0395)		(0.0391)	(0.0381)	(0.0388)	(0.0852)
N_s: Closest tie migrant		0.0499*	0.0442*	0.0662**	–0.0101	–0.0097
		(0.0278)	(0.0268)	(0.0314)	(0.0352)	(0.0361)
$N_w \times N_s$					0.2845**	0.2761**
					(0.1362)	(0.1379)
N_w^2: Village outmigrants sq.						–0.2920***
						(0.1123)
C_w: Village population (/1000)				–0.0031	–0.0027	–0.0023
				(0.0025)	(0.0024)	(0.0024)
C_w: Village self-employed				0.0748	0.0784	0.0792
				(0.3001)	(0.2989)	(0.2957)
C_w: Village average wage (RMB)				–0.0001**	–0.0001**	–0.0001**
				(0.0000)	(0.0000)	(0.0000)
C_s: Relative				–0.0031	–0.0017	–0.0018
				(0.0117)	(0.0117)	(0.0116)
C_s: High Educ.				–0.0269*	–0.0243*	–0.0256*
				(0.0141)	(0.0144)	(0.0141)
C_s: Laborer				0.0166	0.0183	0.0166
				(0.0123)	(0.0123)	(0.0122)
δ: Rate of village immmigration	–0.1558*	–0.1859*	–0.1524*	–0.1035	–0.1112	–0.1042
	(0.0915)	(0.1023)	(0.0917)	(0.0837)	(0.0838)	(0.0821)
λ: Change in urban unempl. rate	–0.2634***	–0.3193***	–0.2622***	–0.2955***	–0.2911***	–0.2798***
	(0.0590)	(0.0588)	(0.0587)	(0.0585)	(0.0589)	(0.0586)
Male (D)	0.0406***	0.0394***	0.0408***	0.0388***	0.0396***	0.0404***
	(0.0122)	(0.0122)	(0.0122)	(0.0118)	(0.0118)	(0.0118)
Age	–0.0081***	–0.0080***	–0.0080***	–0.0076***	–0.0077***	–0.0077***
	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Marital status (D)	0.0021	0.0031	0.0034	0.0080	0.0087	0.0081
	(0.0188)	(0.0188)	(0.0186)	(0.0181)	(0.0180)	(0.0179)
Years of education	–0.0007	–0.0008	–0.0005	–0.0001	–0.0000	–0.0001
	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0022)
Eldest child (D)	–0.0041	–0.0059	–0.0047	–0.0047	–0.0048	–0.0053
	(0.0112)	(0.0112)	(0.0112)	(0.0109)	(0.0109)	(0.0108)
Farmer worker (D)	0.0309**	0.0337***	0.0315**	0.0214*	0.0193	0.0174
	(0.0127)	(0.0128)	(0.0127)	(0.0124)	(0.0125)	(0.0124)
Household size	0.0166***	0.0181***	0.0166***	0.0157***	0.0160***	0.0157***
	(0.0053)	(0.0054)	(0.0053)	(0.0053)	(0.0053)	(0.0052)
Household's number of children	–0.0085	–0.0086	–0.0091	–0.0126	–0.0130	–0.0126
	(0.0115)	(0.0118)	(0.0116)	(0.0110)	(0.0110)	(0.0109)
Household income	–0.0254**	–0.0219*	–0.0238**	–0.0183	–0.0194*	–0.0192*
	(0.0115)	(0.0114)	(0.0115)	(0.0112)	(0.0112)	(0.0112)
Housing value (log RMB)	–0.0224***	–0.0243***	–0.0223***	–0.0198***	–0.0206***	–0.0205***
	(0.0059)	(0.0059)	(0.0058)	(0.0058)	(0.0058)	(0.0057)
Size of family farm land (Mu)	0.0392***	0.0414***	0.0387***	0.0345***	0.0338***	0.0340***
	(0.0084)	(0.0085)	(0.0083)	(0.0083)	(0.0083)	(0.0082)
Emigration region (D)	0.0243*	0.0342**	0.0269*	0.0175	0.0148	0.0132
	(0.0141)	(0.0144)	(0.0142)	(0.0148)	(0.0150)	(0.0149)
Pseudo R^2	0.18	0.17	0.18	0.18	0.19	0.19
N	2192	2192	2192	2192	2192	2192

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
Sample is composed by individuals aged 16 to 35 who have never migrated before.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Weak ties only
Col II: Strong ties only
Col III: Weak and strong ties
Col IV: Weak and strong ties with ties controls
Col V: Weak and strong ties interaction
Col VI: Quadratic weak ties and strong ties interaction

Table 4: Predicted probabilities

	Main effects (Tab 3 Col IV)				Interaction model (Tab 3 Col V)			
	N_s=0	s.e.	N_s=1	s.e.	N_s=0	s.e.	N_s=1	s.e.
N_w=0	0.0865	0.0091	0.1477	0.0296	0.0911	0.0095	0.0811	0.0343
N_w=0.1	0.1021	0.0074	0.1700	0.0305	0.1048	0.0076	0.1255	0.0344
N_w=0.2	0.1195	0.0068	0.1942	0.0321	0.1197	0.0068	0.1840	0.0337
N_w=0.3	0.1390	0.0088	0.2204	0.0348	0.1362	0.0088	0.2566	0.0417
N_w=0.4	0.1604	0.0134	0.2484	0.0386	0.1540	0.0134	0.3411	0.0628
N_w=0.5	0.1839	0.0196	0.2781	0.0436	0.1733	0.0194	0.4337	0.0908
N_w=0.6	0.2092	0.0271	0.3094	0.0498	0.1940	0.0266	0.5293	0.1180

Source: RUMiC data, RHS waves 2008, 2009 and 2010.
Predicted probabilites calculated using values of N_s and N_w from col IV and V of table 3 and keeping other predictors at the mean values.
N_s / N_w is the share of strong / weak ties who migrated.
Strong ties refer to closest person in network; Weak ties refer to village.

Table 5: Excluding household heads

	I	II	III	IV	V	VI
N_w: Village outmigrants	0.1938***		0.1892***	0.1592***	0.1302***	0.3208***
	(0.0430)		(0.0426)	(0.0414)	(0.0420)	(0.0928)
N_s: Closest tie migrant		0.0515*	0.0438	0.0744**	–0.0225	–0.0230
		(0.0298)	(0.0283)	(0.0342)	(0.0341)	(0.0345)
$N_w \times N_s$					0.3666**	0.3594**
					(0.1469)	(0.1483)
N_w^2: Village outmigrants sq.						–0.3029**
						(0.1216)
C_w: Village population (/1000)				–0.0039	–0.0035	–0.0031
				(0.0028)	(0.0027)	(0.0027)
C_w: Village self-employed				0.1143	0.1246	0.1257
				(0.3186)	(0.3155)	(0.3125)
C_w: Village average wage (RMB)				–0.0001***	–0.0001***	–0.0001**
				(0.0000)	(0.0000)	(0.0000)
C_s: Relative				–0.0079	–0.0062	–0.0062
				(0.0127)	(0.0127)	(0.0126)
C_s: High Educ.				–0.0318**	–0.0286*	–0.0300*
				(0.0155)	(0.0159)	(0.0156)
C_s: Laborer				0.0241*	0.0263*	0.0244*
				(0.0136)	(0.0136)	(0.0135)
δ: Rate of village immmigration	–0.1703*	–0.2040*	–0.1674*	–0.1081	–0.1180	–0.1099
	(0.1001)	(0.1111)	(0.1003)	(0.0904)	(0.0902)	(0.0888)
λ: Change in urban unempl. rate	–0.2633***	–0.3255***	–0.2632***	–0.3025***	–0.2950***	–0.2834***
	(0.0649)	(0.0647)	(0.0647)	(0.0643)	(0.0646)	(0.0641)
Male (D)	0.0484***	0.0478***	0.0487***	0.0457***	0.0463***	0.0475***
	(0.0145)	(0.0145)	(0.0145)	(0.0139)	(0.0139)	(0.0139)
Age	–0.0088***	–0.0088***	–0.0088***	–0.0084***	–0.0084***	–0.0084***
	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Marital status (D)	0.0108	0.0125	0.0124	0.0188	0.0196	0.0188
	(0.0205)	(0.0205)	(0.0204)	(0.0197)	(0.0196)	(0.0195)
Years of education	0.0002	–0.0000	0.0003	0.0008	0.0009	0.0009
	(0.0024)	(0.0025)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Eldest child (D)	–0.0088	–0.0109	–0.0095	–0.0096	–0.0096	–0.0100
	(0.0123)	(0.0124)	(0.0123)	(0.0119)	(0.0119)	(0.0118)
Farmer worker (D)	0.0306**	0.0342**	0.0313**	0.0185	0.0154	0.0131
	(0.0139)	(0.0141)	(0.0139)	(0.0136)	(0.0136)	(0.0135)
Household size	0.0160***	0.0176***	0.0160***	0.0151**	0.0155***	0.0151**
	(0.0060)	(0.0062)	(0.0061)	(0.0060)	(0.0060)	(0.0059)
Household's number of children	–0.0106	–0.0108	–0.0115	–0.0165	–0.0170	–0.0165
	(0.0136)	(0.0138)	(0.0137)	(0.0131)	(0.0131)	(0.0130)
Household income	–0.0294**	–0.0254**	–0.0277**	–0.0216*	–0.0231*	–0.0230*
	(0.0128)	(0.0127)	(0.0128)	(0.0124)	(0.0124)	(0.0124)
Housing value (log RMB)	–0.0247***	–0.0266***	–0.0246***	–0.0217***	–0.0229***	–0.0228***
	(0.0065)	(0.0065)	(0.0064)	(0.0063)	(0.0063)	(0.0063)
Size of family farm land (Mu)	0.0412***	0.0435***	0.0408***	0.0350***	0.0340***	0.0343***
	(0.0092)	(0.0093)	(0.0092)	(0.0091)	(0.0091)	(0.0090)
Emigration region (D)	0.0227	0.0335**	0.0254	0.0143	0.0104	0.0087
	(0.0154)	(0.0158)	(0.0156)	(0.0161)	(0.0162)	(0.0162)
Pseudo R^2	0.17	0.16	0.17	0.18	0.18	0.19
N	2005	2005	2005	2005	2005	2005

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 35 who have never migrated before and are not household heads.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Weak ties only
Col II: Strong ties only
Col III: Weak and strong ties
Col IV: Weak and strong ties with ties controls
Col V: Weak and strong ties interaction
Col VI: Quadratic weak ties and strong ties interaction

Table 6: Knows closest tie since before age 16

	I	II	III	IV	V	VI
N_w: Village outmigrants	0.2090***		0.2044***	0.1771***	0.1482***	0.3878***
	(0.0474)		(0.0468)	(0.0463)	(0.0470)	(0.1052)
N_s: Closest tie migrant		0.0679**	0.0621*	0.0917**	–0.0059	–0.0050
		(0.0338)	(0.0329)	(0.0388)	(0.0446)	(0.0457)
$N_w \times N_s$					0.3617**	0.3501**
					(0.1675)	(0.1683)
N_w^2: Village outmigrants sq.						–0.3794***
						(0.1379)
C_w: Village population (/1000)				–0.0031	–0.0027	–0.0022
				(0.0029)	(0.0028)	(0.0028)
C_w: Village self-employed				0.2568	0.2610	0.2663
				(0.3627)	(0.3611)	(0.3577)
C_w: Village average wage (RMB)				–0.0001**	–0.0001**	–0.0001**
				(0.0000)	(0.0000)	(0.0000)
C_s: Relative				–0.0106	–0.0091	–0.0096
				(0.0142)	(0.0142)	(0.0141)
C_s: High Educ.				–0.0284	–0.0254	–0.0276
				(0.0184)	(0.0187)	(0.0184)
C_s: Laborer				0.0204	0.0224	0.0199
				(0.0154)	(0.0153)	(0.0152)
δ: Rate of village immmigration	–0.1738	–0.2060*	–0.1698	–0.1257	–0.1364	–0.1268
	(0.1127)	(0.1251)	(0.1134)	(0.1069)	(0.1069)	(0.1049)
λ: Change in urban unempl. rate	–0.3162***	–0.3825***	–0.3126***	–0.3511***	–0.3435***	–0.3279***
	(0.0727)	(0.0715)	(0.0722)	(0.0722)	(0.0728)	(0.0723)
Male (D)	0.0489***	0.0483***	0.0493***	0.0476***	0.0486***	0.0499***
	(0.0148)	(0.0148)	(0.0148)	(0.0145)	(0.0145)	(0.0145)
Age	–0.0085***	–0.0084***	–0.0084***	–0.0081***	–0.0082***	–0.0082***
	(0.0020)	(0.0020)	(0.0020)	(0.0019)	(0.0019)	(0.0019)
Marital status (D)	0.0048	0.0055	0.0065	0.0112	0.0123	0.0116
	(0.0228)	(0.0227)	(0.0227)	(0.0223)	(0.0222)	(0.0220)
Years of education	–0.0021	–0.0024	–0.0020	–0.0016	–0.0015	–0.0016
	(0.0028)	(0.0028)	(0.0028)	(0.0027)	(0.0028)	(0.0027)
Eldest child (D)	–0.0051	–0.0069	–0.0057	–0.0051	–0.0049	–0.0055
	(0.0138)	(0.0138)	(0.0138)	(0.0135)	(0.0135)	(0.0134)
Farmer worker (D)	0.0324**	0.0365**	0.0335**	0.0224	0.0195	0.0168
	(0.0154)	(0.0155)	(0.0154)	(0.0154)	(0.0155)	(0.0155)
Household size	0.0227***	0.0246***	0.0228***	0.0223***	0.0227***	0.0223***
	(0.0066)	(0.0066)	(0.0065)	(0.0066)	(0.0066)	(0.0065)
Household's number of children	–0.0101	–0.0092	–0.0109	–0.0140	–0.0146	–0.0141
	(0.0142)	(0.0145)	(0.0144)	(0.0137)	(0.0138)	(0.0136)
Household income	–0.0242*	–0.0197	–0.0219	–0.0157	–0.0174	–0.0169
	(0.0141)	(0.0140)	(0.0141)	(0.0138)	(0.0139)	(0.0140)
Housing value (log RMB)	–0.0263***	–0.0284***	–0.0262***	–0.0243***	–0.0253***	–0.0253***
	(0.0071)	(0.0071)	(0.0070)	(0.0071)	(0.0071)	(0.0070)
Size of family farm land (Mu)	0.0465***	0.0484***	0.0457***	0.0409***	0.0402***	0.0407***
	(0.0103)	(0.0103)	(0.0102)	(0.0103)	(0.0103)	(0.0102)
Emigration region (D)	0.0333*	0.0453**	0.0369**	0.0279	0.0242	0.0216
	(0.0175)	(0.0176)	(0.0175)	(0.0184)	(0.0186)	(0.0186)
Pseudo R^2	0.16	0.15	0.16	0.17	0.17	0.18
N	1842	1842	1842	1842	1842	1842

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 35 who have never migrated before and who know their closest tie since before age 16.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Weak ties only
Col II: Strong ties only
Col III: Weak and strong ties
Col IV: Weak and strong ties with ties controls
Col V: Weak and strong ties interaction
Col VI: Quadratic weak ties and strong ties interaction

Table 7: Excluding household heads and knows closest tie since before age 16

	I	II	III	IV	V	VI
N_w: Village outmigrants	0.2157***		0.2105***	0.1785***	0.1428***	0.3928***
	(0.0502)		(0.0496)	(0.0490)	(0.0498)	(0.1121)
N_s: Closest tie migrant		0.0650*	0.0576*	0.0947**	–0.0239	–0.0242
		(0.0349)	(0.0335)	(0.0404)	(0.0412)	(0.0417)
$N_w \times N_s$					0.4570***	0.4465**
					(0.1772)	(0.1773)
N_w^2: Village outmigrants sq.						–0.3914***
						(0.1468)
C_w: Village population (/1000)				–0.0039	–0.0035	–0.0030
				(0.0032)	(0.0031)	(0.0031)
C_w: Village self-employed				0.3078	0.3191	0.3273
				(0.3773)	(0.3738)	(0.3705)
C_w: Village average wage (RMB)				–0.0001**	–0.0001**	–0.0001**
				(0.0000)	(0.0000)	(0.0000)
C_s: Relative				–0.0161	–0.0143	–0.0147
				(0.0152)	(0.0152)	(0.0151)
C_s: High Educ.				–0.0343*	–0.0308	–0.0331*
				(0.0195)	(0.0199)	(0.0195)
C_s: Laborer				0.0259	0.0285*	0.0257
				(0.0165)	(0.0164)	(0.0163)
δ: Rate of village immmigration	–0.1835	–0.2188*	–0.1803	–0.1298	–0.1432	–0.1325
	(0.1189)	(0.1311)	(0.1197)	(0.1120)	(0.1115)	(0.1098)
λ: Change in urban unempl. rate	–0.3179***	–0.3887***	–0.3160***	–0.3592***	–0.3477***	–0.3321***
	(0.0776)	(0.0766)	(0.0773)	(0.0773)	(0.0778)	(0.0772)
Male (D)	0.0570***	0.0567***	0.0574***	0.0548***	0.0557***	0.0575***
	(0.0169)	(0.0169)	(0.0169)	(0.0164)	(0.0164)	(0.0164)
Age	–0.0091***	–0.0092***	–0.0091***	–0.0088***	–0.0089***	–0.0089***
	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)
Marital status (D)	0.0109	0.0124	0.0130	0.0194	0.0204	0.0196
	(0.0247)	(0.0245)	(0.0246)	(0.0241)	(0.0240)	(0.0238)
Years of education	–0.0008	–0.0012	–0.0007	–0.0003	–0.0002	–0.0002
	(0.0029)	(0.0030)	(0.0030)	(0.0029)	(0.0029)	(0.0029)
Eldest child (D)	–0.0093	–0.0113	–0.0099	–0.0093	–0.0089	–0.0095
	(0.0148)	(0.0149)	(0.0148)	(0.0144)	(0.0144)	(0.0143)
Farmer worker (D)	0.0318*	0.0365**	0.0329**	0.0197	0.0155	0.0124
	(0.0166)	(0.0167)	(0.0166)	(0.0164)	(0.0165)	(0.0164)
Household size	0.0219***	0.0238***	0.0220***	0.0215***	0.0221***	0.0216***
	(0.0072)	(0.0073)	(0.0072)	(0.0072)	(0.0072)	(0.0071)
Household's number of children	–0.0082	–0.0077	–0.0094	–0.0140	–0.0145	–0.0140
	(0.0161)	(0.0163)	(0.0163)	(0.0156)	(0.0157)	(0.0155)
Household income	–0.0296*	–0.0248	–0.0274*	–0.0207	–0.0231	–0.0227
	(0.0153)	(0.0151)	(0.0153)	(0.0150)	(0.0150)	(0.0151)
Housing value (log RMB)	–0.0292***	–0.0312***	–0.0290***	–0.0269***	–0.0282***	–0.0283***
	(0.0077)	(0.0077)	(0.0076)	(0.0076)	(0.0076)	(0.0076)
Size of family farm land (Mu)	0.0479***	0.0499***	0.0473***	0.0410***	0.0400***	0.0407***
	(0.0109)	(0.0110)	(0.0109)	(0.0109)	(0.0109)	(0.0109)
Emigration region (D)	0.0308*	0.0436**	0.0346*	0.0241	0.0188	0.0161
	(0.0187)	(0.0189)	(0.0188)	(0.0196)	(0.0198)	(0.0197)
Pseudo R^2	0.16	0.15	0.16	0.17	0.17	0.18
N	1717	1717	1717	1717	1717	1717

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 35 who have never migrated before, are not household heads and who know their closest tie since before age 16..
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Weak ties only
Col II: Strong ties only
Col III: Weak and strong ties
Col IV: Weak and strong ties with ties controls
Col V: Weak and strong ties interaction
Col VI: Quadratic weak ties and strong ties interaction

Table 8: All individuals aged 16 to 64

	I	II	III	IV	V	VI
N_w: Village outmigrants	0.0258***		0.0254***	0.0212***	0.0145*	0.0514***
	(0.0073)		(0.0073)	(0.0071)	(0.0074)	(0.0181)
N_s: Closest tie migrant		0.0047	0.0041	0.0079	–0.0073	–0.0085*
		(0.0043)	(0.0041)	(0.0049)	(0.0047)	(0.0046)
$N_w \times N_s$					0.0622***	0.0688***
					(0.0214)	(0.0233)
N_w^2: Village outmigrants sq.						–0.0619**
						(0.0262)
C_w: Village population (/1000)				–0.0017**	–0.0016**	–0.0016**
				(0.0007)	(0.0007)	(0.0007)
C_w: Village self-employed				–0.0562	–0.0562	–0.0566
				(0.0629)	(0.0629)	(0.0624)
C_w: Village average wage (RMB)				–0.0000**	–0.0000***	–0.0000**
				(0.0000)	(0.0000)	(0.0000)
C_s: Relative				0.0003	0.0004	0.0004
				(0.0022)	(0.0022)	(0.0022)
C_s: High Educ.				–0.0089***	–0.0084***	–0.0083***
				(0.0025)	(0.0025)	(0.0025)
C_s: Laborer				0.0022	0.0025	0.0023
				(0.0023)	(0.0023)	(0.0023)
δ: Rate of village immmigration	–0.0627***	–0.0691***	–0.0623***	–0.0455**	–0.0471**	–0.0447**
	(0.0229)	(0.0243)	(0.0229)	(0.0204)	(0.0204)	(0.0199)
λ: Change in urban unempl. rate	–0.0524***	–0.0594***	–0.0522***	–0.0562***	–0.0551***	–0.0526***
	(0.0106)	(0.0105)	(0.0106)	(0.0104)	(0.0104)	(0.0103)
Male (D)	0.0123***	0.0122***	0.0124***	0.0117***	0.0117***	0.0118***
	(0.0025)	(0.0026)	(0.0025)	(0.0025)	(0.0025)	(0.0024)
Age	–0.0019***	–0.0019***	–0.0019***	–0.0018***	–0.0018***	–0.0018***
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Marital status (D)	–0.0059	–0.0058	–0.0058	–0.0046	–0.0044	–0.0047
	(0.0043)	(0.0042)	(0.0042)	(0.0040)	(0.0040)	(0.0040)
Years of education	–0.0004	–0.0004	–0.0004	–0.0003	–0.0003	–0.0003
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Eldest child (D)	–0.0012	–0.0013	–0.0012	–0.0010	–0.0009	–0.0010
	(0.0022)	(0.0022)	(0.0022)	(0.0021)	(0.0021)	(0.0021)
Farmer worker (D)	0.0041*	0.0044*	0.0041*	0.0020	0.0017	0.0014
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Household size	0.0045***	0.0048***	0.0045***	0.0043***	0.0042***	0.0042***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Household's number of children	–0.0005	–0.0004	–0.0005	–0.0010	–0.0010	–0.0009
	(0.0015)	(0.0015)	(0.0015)	(0.0014)	(0.0014)	(0.0014)
Household income	–0.0061***	–0.0059***	–0.0061***	–0.0055***	–0.0056***	–0.0056***
	(0.0014)	(0.0014)	(0.0014)	(0.0016)	(0.0016)	(0.0016)
Housing value (log RMB)	–0.0041***	–0.0044***	–0.0041***	–0.0033***	–0.0035***	–0.0034***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
Size of family farm land (Mu)	0.0045***	0.0047***	0.0044***	0.0037**	0.0035**	0.0036**
	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Emigration region (D)	0.0080***	0.0091***	0.0081***	0.0050*	0.0046*	0.0044*
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
Pseudo R^2	0.20	0.19	0.20	0.20	0.20	0.21
N	11777	11777	11777	11777	11777	11777

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 64 who have never migrated before.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Weak ties only
Col II: Strong ties only
Col III: Weak and strong ties
Col IV: Weak and strong ties with ties controls
Col V: Weak and strong ties interaction
Col VI: Quadratic weak ties and strong ties interaction

Table 9: Including individuals who migrated before

	I	II	III	IV	V	VI
N_w: Village outmigrants	0.4722***		0.4668***	0.4127***	0.3777***	1.1667***
	(0.0594)		(0.0591)	(0.0592)	(0.0610)	(0.1360)
N_s: Closest tie migrant		0.1048***	0.1002***	0.1252***	0.0194	0.0361
		(0.0233)	(0.0233)	(0.0249)	(0.0531)	(0.0513)
$N_w \times N_s$					0.4433**	0.3557*
					(0.2066)	(0.1934)
N_w^2: Village outmigrants sq.						–1.2544***
						(0.1995)
C_w: Village population (/1000)				0.0033	0.0034	0.0053
				(0.0035)	(0.0035)	(0.0035)
C_w: Village self-employed				0.2617	0.2628	0.2830
				(0.3893)	(0.3888)	(0.3930)
C_w: Village average wage (RMB)				–0.0004***	–0.0004***	–0.0004***
				(0.0001)	(0.0001)	(0.0001)
C_s: Relative				0.0405***	0.0412***	0.0390***
				(0.0145)	(0.0145)	(0.0145)
C_s: High Educ.				–0.0658***	–0.0629***	–0.0665***
				(0.0204)	(0.0205)	(0.0205)
C_s: Laborer				0.0466***	0.0473***	0.0443***
				(0.0154)	(0.0154)	(0.0154)
δ: Rate of village immmigration	–0.8337***	–0.9596***	–0.8221***	–0.6058***	–0.6179***	–0.5712***
	(0.1621)	(0.1787)	(0.1621)	(0.1497)	(0.1508)	(0.1486)
λ: Change in urban unempl. rate	–0.1672**	–0.2816***	–0.1699**	–0.3099***	–0.3057***	–0.2828***
	(0.0691)	(0.0684)	(0.0693)	(0.0714)	(0.0715)	(0.0718)
Male (D)	0.1098***	0.1107***	0.1115***	0.1109***	0.1117***	0.1148***
	(0.0140)	(0.0140)	(0.0141)	(0.0141)	(0.0141)	(0.0142)
Age	–0.0097***	–0.0102***	–0.0099***	–0.0092***	–0.0093***	–0.0095***
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Marital status (D)	–0.0848***	–0.0758***	–0.0833***	–0.0777***	–0.0779***	–0.0813***
	(0.0231)	(0.0230)	(0.0231)	(0.0232)	(0.0232)	(0.0232)
Years of education	–0.0061**	–0.0071**	–0.0064**	–0.0038	–0.0037	–0.0041
	(0.0031)	(0.0030)	(0.0031)	(0.0031)	(0.0031)	(0.0031)
Eldest child (D)	–0.0207	–0.0227	–0.0196	–0.0180	–0.0178	–0.0192
	(0.0140)	(0.0139)	(0.0140)	(0.0140)	(0.0140)	(0.0141)
Farmer worker (D)	–0.0910***	–0.0924***	–0.0906***	–0.1013***	–0.1026***	–0.1000***
	(0.0176)	(0.0176)	(0.0177)	(0.0177)	(0.0177)	(0.0178)
Household size	0.0550***	0.0580***	0.0549***	0.0484***	0.0485***	0.0477***
	(0.0070)	(0.0070)	(0.0070)	(0.0071)	(0.0071)	(0.0071)
Household's number of children	0.0346**	0.0330**	0.0346**	0.0258*	0.0263*	0.0247*
	(0.0153)	(0.0152)	(0.0153)	(0.0149)	(0.0149)	(0.0149)
Household income	–0.0331**	–0.0329**	–0.0307**	–0.0059	–0.0076	–0.0050
	(0.0136)	(0.0138)	(0.0137)	(0.0136)	(0.0136)	(0.0136)
Housing value (log RMB)	–0.0342***	–0.0391***	–0.0350***	–0.0220***	–0.0223***	–0.0218***
	(0.0077)	(0.0076)	(0.0076)	(0.0078)	(0.0078)	(0.0078)
Size of family farm land (Mu)	0.0602***	0.0642***	0.0605***	0.0539***	0.0530***	0.0537***
	(0.0110)	(0.0110)	(0.0110)	(0.0111)	(0.0111)	(0.0112)
Emigration region (D)	0.0373**	0.0508***	0.0408**	0.0117	0.0094	0.0086
	(0.0170)	(0.0168)	(0.0170)	(0.0177)	(0.0178)	(0.0178)
Migrated before	0.4349***	0.4456***	0.4331***	0.4224***	0.4223***	0.4155***
	(0.0129)	(0.0126)	(0.0130)	(0.0134)	(0.0134)	(0.0136)
Pseudo R^2	0.26	0.25	0.26	0.27	0.27	0.28
N	6584	6584	6584	6584	6584	6584

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 35 who have never migrated before and are not household heads.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Weak ties only
Col II: Strong ties only
Col III: Weak and strong ties
Col IV: Weak and strong ties with ties controls
Col V: Weak and strong ties interaction
Col VI: Quadratic weak ties and strong ties interaction

Table 10: Three closest ties

	I	II	III	IV	V
N_w: Village outmigrants	0.1473***	0.1284***	0.1054**	0.2400***	0.2406***
	(0.0415)	(0.0406)	(0.0414)	(0.0912)	(0.0912)
$\overline{N_s}$: Three closest ties migrant (share)	0.0752**	0.1004***	0.0186	0.0191	0.0014
	(0.0307)	(0.0312)	(0.0695)	(0.0702)	(0.1054)
$\overline{N_s} \times N_w$			0.3298	0.3184	0.3182
			(0.2184)	(0.2194)	(0.2200)
$\overline{N_s^2}$					0.0216
					(0.1224)
N_w^2: Village outmigrants sq.				–0.2158*	–0.2166*
				(0.1172)	(0.1171)
C_w: Village population (/1000)		–0.0032	–0.0030	–0.0028	–0.0028
		(0.0030)	(0.0028)	(0.0029)	(0.0029)
C_w: Village self-employed		0.0027	0.0137	0.0201	0.0168
		(0.3421)	(0.3414)	(0.3391)	(0.3401)
C_w: Village average wage (RMB)		–0.0001	–0.0001	–0.0001	–0.0001
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
C_s: Relative		0.0076	0.0096	0.0099	0.0099
		(0.0130)	(0.0131)	(0.0131)	(0.0131)
C_s: High Educ.		–0.0326**	–0.0301**	–0.0311**	–0.0310**
		(0.0148)	(0.0152)	(0.0149)	(0.0150)
C_s: Laborer		0.0301**	0.0316**	0.0302**	0.0301**
		(0.0139)	(0.0140)	(0.0139)	(0.0139)
Pseudo R^2	0.19	0.20	0.20	0.20	0.20
N	1719	1719	1719	1719	1719

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 35 who have never migrated before and are not household heads.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
$\overline{N_s}$ / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Weak and strong ties
Col II: Weak and strong ties with ties controls
Col III: Weak and strong ties interaction
Col IV: Quadratic weak ties and strong ties interaction
Col V: Quadratic weak and strong ties and weak/strong ties interaction

Table 11: Interaction with tie characteristics

	I	II	III	IV
N_w: Village outmigrants	0.1117**	0.1127**	0.0628	0.2692**
	(0.0437)	(0.0441)	(0.0607)	(0.1122)
N_s: Closest tie migrant	–0.0086	–0.0092	–0.0166	–0.0185
	(0.0358)	(0.0352)	(0.0332)	(0.0329)
$N_w \times N_s$	0.2809**	0.2833**	0.3270**	0.3275**
	(0.1361)	(0.1356)	(0.1412)	(0.1438)
$N_w \times N_s$: Relative	0.0674	0.0683	0.0723	0.0175
	(0.0809)	(0.0816)	(0.0820)	(0.0878)
$N_w \times N_s$: High Educ.		–0.0203	–0.0048	–0.0586
		(0.1534)	(0.1530)	(0.1593)
$N_w \times N_s$: Laborer			0.0853	0.0793
			(0.0744)	(0.0795)
N_w^2: Village outmigrants sq.				–0.3000**
				(0.1181)
C_w: Village population (/1000)	–0.0028	–0.0028	–0.0028	–0.0024
	(0.0024)	(0.0024)	(0.0025)	(0.0025)
C_w: Village self-employed	0.0741	0.0737	0.0685	0.0719
	(0.2985)	(0.2987)	(0.2956)	(0.2941)
C_w: Village average wage (RMB)	–0.0001**	–0.0001**	–0.0001**	–0.0001**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
C_s: Relative	–0.0146	–0.0147	–0.0151	–0.0045
	(0.0192)	(0.0193)	(0.0192)	(0.0210)
C_s: High Educ.	–0.0248*	–0.0216	–0.0241	–0.0162
	(0.0143)	(0.0287)	(0.0279)	(0.0308)
C_s: Laborer	0.0184	0.0184	0.0010	0.0005
	(0.0123)	(0.0123)	(0.0196)	(0.0207)
Pseudo R^2	0.19	0.19	0.19	0.19
N	2192	2192	2192	2192

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 35 who have never migrated before and are not household heads.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I: Interaction with relative
Col II: Interaction with high education
Col III: Interaction with laborer
Col IV: Quadratic weak ties and strong ties interactions

Table 12: Sensitivity to migration definition

	I	II	III	IV	V	VI	VII	VIII	IX
N_w: Village outmigrants	0.2492***	0.2270***	0.4899***	0.1779***	0.1516***	0.4446***	0.0855***	0.0785***	0.1518***
	(0.0502)	(0.0579)	(0.1103)	(0.0440)	(0.0455)	(0.1007)	(0.0215)	(0.0217)	(0.0514)
N_s: Closest tie migrant	0.0206	0.0082	0.0053	0.0739**	0.0180	0.0238	0.0659*	0.0032	0.0030
	(0.0166)	(0.0260)	(0.0261)	(0.0291)	(0.0373)	(0.0395)	(0.0393)	(0.0394)	(0.0399)
$N_w \times N_s$		0.0618	0.0859		0.2157*	0.1890		0.1243	0.1266
		(0.1054)	(0.1024)		(0.1264)	(0.1287)		(0.0995)	(0.1003)
N_w^2: Village outmigrants sq.			−0.4418***			−0.4782***			−0.1190*
			(0.1469)			(0.1397)			(0.0703)
C_w: Village population (/1000)	−0.0100**	−0.0099**	−0.0093**	−0.0053*	−0.0049*	−0.0043	−0.0007	−0.0007	−0.0005
	(0.0042)	(0.0042)	(0.0042)	(0.0030)	(0.0029)	(0.0029)	(0.0013)	(0.0013)	(0.0013)
C_w: Village self-employed	−0.3570	−0.3741	−0.3742	−0.1418	−0.1588	−0.1510	0.0876	0.0833	0.0880
	(0.3983)	(0.3973)	(0.3936)	(0.3607)	(0.3623)	(0.3566)	(0.1533)	(0.1535)	(0.1517)
C_w: Village average wage (RMB)	−0.0001**	−0.0001**	−0.0001**	−0.0001***	−0.0001**	−0.0001**	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
C_s: Relative	0.0099	0.0101	0.0091	0.0041	0.0043	0.0039	−0.0014	−0.0006	−0.0004
	(0.0153)	(0.0153)	(0.0152)	(0.0131)	(0.0131)	(0.0129)	(0.0067)	(0.0068)	(0.0067)
C_s: High Educ.	−0.0299*	−0.0294	−0.0321*	−0.0378**	−0.0358**	−0.0382**	−0.0055	−0.0042	−0.0050
	(0.0181)	(0.0181)	(0.0178)	(0.0153)	(0.0156)	(0.0151)	(0.0088)	(0.0091)	(0.0088)
C_s: Laborer	0.0083	0.0090	0.0079	0.0167	0.0172	0.0150	0.0009	0.0014	0.0010
	(0.0151)	(0.0151)	(0.0150)	(0.0138)	(0.0138)	(0.0136)	(0.0068)	(0.0069)	(0.0068)
Pseudo R^2	0.17	0.17	0.18	0.18	0.18	0.19	0.25	0.25	0.25
N	2192	2192	2192	2192	2192	2192	2192	2192	2192

Source: RUMiC data, RHS waves 2008, 2009 and 2010. Robust standard errors in parenthesis.
Sample is composed by individuals aged 16 to 35 who have never migrated before and are not household heads.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
(D) refers to dummy variables.
N_s / N_w is the share of strong / weak ties who migrated.
C_s / C_w refer to other characteristics of the strong / weak ties.
Strong ties refer to closest person in network; Weak ties refer to village.
Col I-III: Migrated within rural or urban areas
Col IV-VI: Migrated urban areas within county
Col VII-IX: Migrated urban areas another province

Figure 2: Strong ties: Type of help

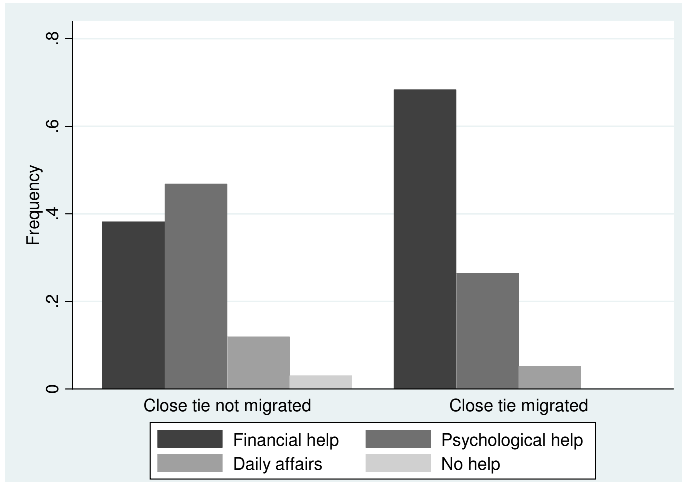

Figure 3: Strong ties: Frequency of contacts

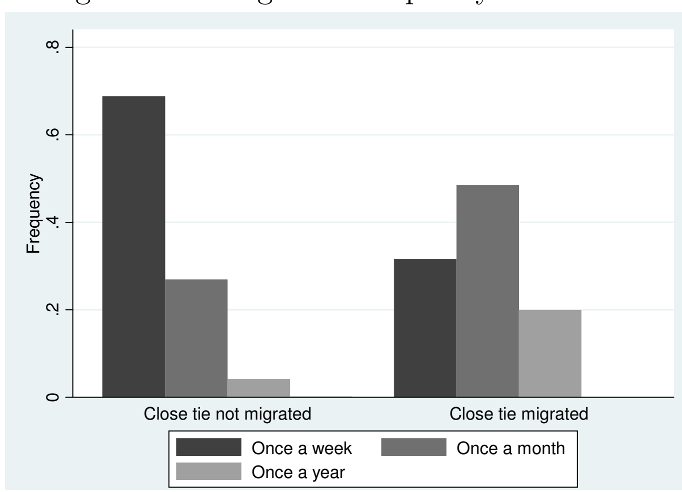

Figure 4: Strong ties: Money/gifts exchange

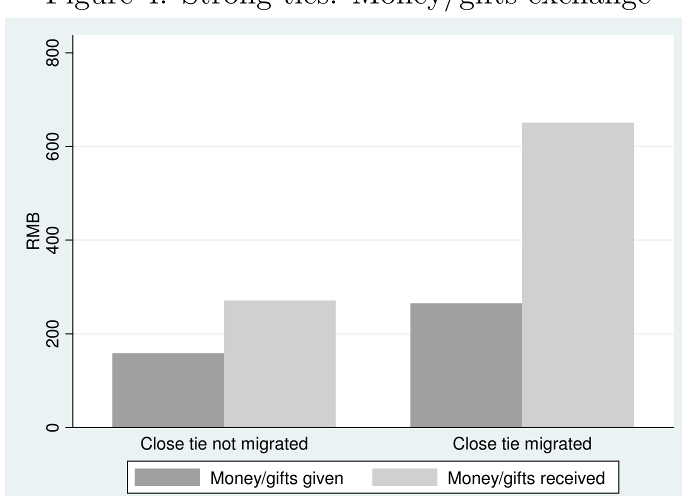

Figure 5: Predicted probabilities: main effects

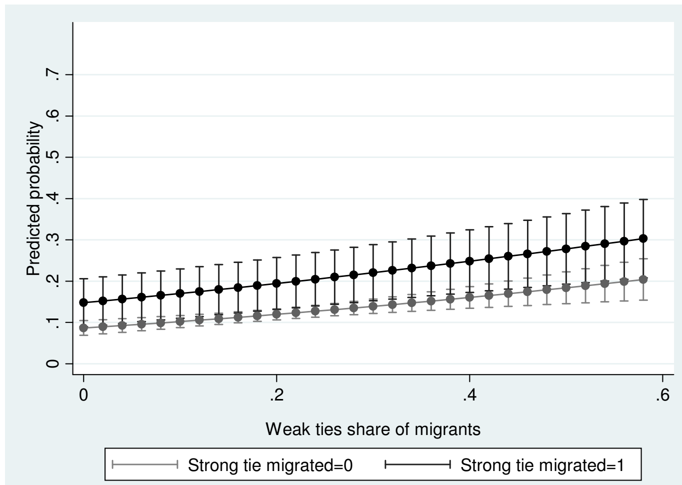

Figure 6: Predicted probabilities: interaction