Review of Economic Studies (2023) **00**, 1–32 doi:10.1093/restud/rdad113

Advance access publication 14 December 2023

Migration and the Value of Social Networks

Joshua E. Blumenstock and Guanghua Chi
University of California, USA

and

Xu Tan
University of Washington, USA

First version received September 2019*; Editorial decision June* 2023*; Accepted December* 2023 *(Eds.)*

How do social networks influence the decision to migrate? Prior work suggests two distinct mechanisms that have historically been difficult to differentiate: as a conduit of information, and as a source of social and economic support. We disentangle these mechanisms using a massive "digital trace" dataset that allows us to observe the migration decisions made by millions of individuals over several years, as well as the complete social network of each person in the months before and after migration. These data allow us to establish a new set of stylized facts about the relationship between social networks and migration. Our main analysis indicates that the average migrant derives more social capital from "interconnected" networks that provide social support than from "extensive" networks that efficiently transmit information.

Key words: Networks, Migration, Social Networks, Social Capital, Big Data, Development

JEL codes: O15, R23, D85, Z13, O12, C55

1. INTRODUCTION

The decision to migrate is one of the most important economic decisions an individual can make. Many factors influence this decision, from employment prospects and amenity differentials to life-cycle considerations and migration costs. In each of these factors, social networks play a prominent role. It is through social networks that migrants learn about opportunities and conditions in potential destinations; at home, the structure of migrants' social networks shapes their ability and desire to leave.

This paper uses a rich source of digital data to add considerable nuance to our understanding of *how* social networks influence an individual's decision to migrate. Here, prior work emphasizes two distinct mechanisms: first, that networks provide migrants with access to information, for instance about jobs and conditions in the destination (Borjas, 1992; Topa, 2001; Munshi, 2003; Dustmann *et al.*, 2016); and second, that networks act as a safety net for migrants by providing material or social support (Carrington *et al.*, 1996; Edin *et al.*, 2003; Dolfin and Genicot, 2010; Munshi, 2014; Comola and Mendola, 2015). However, there is considerable ambiguity about the nature and relative importance of these two mechanisms. For instance,

The editor in charge of this paper was Thomas Chaney.

the prevailing view in the migration literature is that migrants tend to go to places where they have larger networks, but a handful of studies argue that larger networks may actually deter migration, for instance if migrants compete with one another over opportunities and resources.[1] Similarly, robust risk sharing networks can both facilitate migration by providing informal insurance against negative outcomes (Morten, 2019), and discourage migration if migrants fear those left behind will be sanctioned for their departure (Banerjee and Newman, 1998; Munshi and Rosenzweig, 2016).

These ambiguities arise in part because it is difficult to link social network structure to migration decisions using traditional data (Chuang and Schechter, 2015). Instead, most existing work relies on indirect proxies for a migrant's social network, such as the assumption that individuals from the same hometown, or with similar observable characteristics, are more likely to be connected than two dissimilar individuals.[2] Such proxies can provide a reasonable approximation of the size of a migrant's social network, but they do not reveal if and how other aspects of social network structure influence the migration decision. Higher-order network structure—that is, the connections of an individual's connections—plays a critical role in decisions about employment, education, health, finance, product adoption, and the formation of strategic alliances.[3] Yet, the role of such network structure in migration has not been systematically studied.

We leverage a rich new source of "digital trace" data to provide a detailed empirical perspective on how social networks influence the decision to migrate. These data capture the entire universe of mobile phone activity in Rwanda over a 5-year period. Each of roughly 1 million individuals is uniquely identified throughout the dataset, and every time they make or receive a phone call, we observe their approximate location, as well as the identity of the person they are talking to. From these data, we can reconstruct each subscriber's 5-year migration trajectory, as well as a detailed picture of their social network before and after migration.

The empirical analysis links each individual's migration decisions over time to the evolving structure of their social network. For instance, we use these data to confirm the longstanding hypothesis that people move to places where they know more people; conversely, individuals are less likely to leave places where they have larger networks. While these results may be intuitive, our data make it possible to disaggregate this relationship in a way that has not been done previously. In particular, we observe migration decisions for every possible network size and structure; we thus can estimate, for instance, that roughly 4% of individuals with ten contacts in a potential destination d eventually migrate to that location. More broadly, we observe that

1. Classic papers documenting the "prevailing" view include Rees (1966), Greenwood (1969), Granovetter (1973), Montgomery (1991), and Borjas *et al.* (1992). More recent examples include Munshi (2003), Winters *et al.* (2001), Dolfin and Genicot (2010), Patel and Vella (2012), Fafchamps and Shilpi (2013), Mahajan and Yang (2017), Giulietti *et al.* (2018), and Bertoli and Ruyssen (2018). Papers that highlight the potential deterrent effect of larger networks include Calvó-Armengol (2004), Calvó-Armengol and Jackson (2004), Wahba and Zenou (2005), and Beaman (2012).

2. For instance, Munshi (2003) uses rainfall shocks at origin to instrument for network size at destination. Beaman (2012) exploits exogenous variation in the size of the migrant's social network induced by the quasi-random assignment of political refugees to new communities. Kinnan *et al.* (2018) take advantage of a resettlement program in China that sent 18 million urban youth to rural areas. Related approaches are used by Card (2001), Hanson and Woodruff (2003), and Dinkelman and Mariotti (2016).

3. For example: Granovetter (1973), Burt (1992), and Karlan *et al.* (2009) provide examples of how higher-order network structure affects employment prospects. Banerjee *et al.* (2013), Beaman *et al.* (2015), and Ugander *et al.* (2012) illustrate the importance of higher-order structure in the adoption of microfinance, new plant seeds, and Facebook, respectively. Ambrus *et al.* (2015) and Chandrasekhar *et al.* (2018) relate network structure to contract enforcement and informal insurance. Keeling and Eames (2005) review how network structure influences the spread of infectious diseases. König *et al.* (2017) and Jackson and Nei (2015) link political network structure to strategic alliance formation. See Jackson (2010) and Easley and Kleinberg (2010) for an overview.

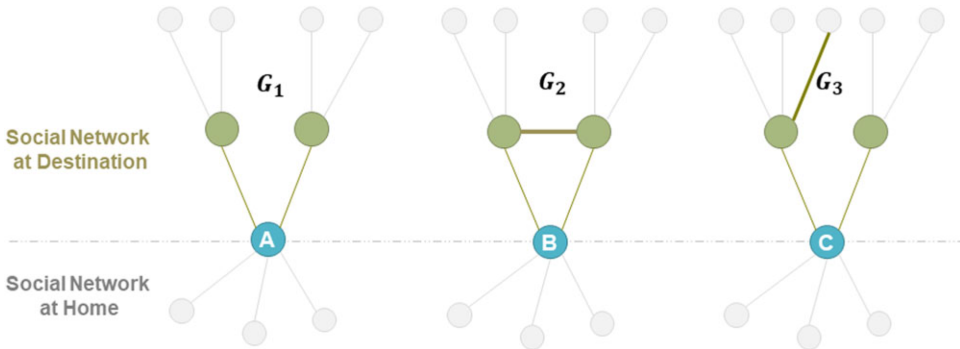

FIGURE 1
Schematic diagrams of the social networks of three migrants

Notes: Each of the blue circles (labeled A, B, C) represents a different individual considering migrating from their home to a new destination. Each individual has exactly three contacts in the home district (smaller grey circles below the dashed line) and two contacts in the destination district (larger green circles above the dashed line). The social network of these three individuals is denoted by G_1, G_2, and G_3.

the relationship between migration and network size is positive, monotonic, and approximately linear with slope of unity, such that the probability of migration roughly doubles as the number of contacts in the destination doubles. Superficially, this result diverges from a series of studies that predict eventual negative externalities from network size, as when members compete for information and opportunities (Calvó-Armengol, 2004; Calvó-Armengol and Jackson, 2004; Beaman, 2012; Dagnelie *et al.*, 2019).

We then focus on developing a systematic understanding of how *higher-order* network structure—that is, the connections of the migrant's connections (and their connections' connections, and so forth)—influences the decision to migrate. This is again something that would be very difficult to study with traditional survey data, but which we observe in rich detail in the mobile phone records. The purpose of this analysis is to understand whether, ceteris paribus, individuals are more likely to migrate to places where their social networks have particular network topologies. A stylized version of our approach is shown in Figure 1: we are interested in understanding whether, for instance, individual A is more likely to migrate than individual B, where both A and B know exactly two people in the destination and three people at home, and the only observable difference between A and B is that B's contacts are connected to each other, whereas A's contacts are from two disjoint communities.

Our ability to identify the effect of social networks on migration is complicated by the fact that network structure is not exogenous. We address this concern in three principal ways. First, as noted, we focus on the relationship between the *higher-order* structure of a migrant's social network and subsequent migration decision. While a migrant may easily influence their direct connections, we assume they have less ability to influence the exact manner in which their connections are connected to one another. Second, we relate migration decisions in each month to the higher-order structure of the network several months prior. This is meant to minimize the likelihood that the decision to migrate shaped the social network, rather than vice versa.[4] Finally, we use an extremely restrictive set of fixed effects to eliminate many likely sources

4. Our main specifications relate migration decisions to network structure 2 months prior, but results are unchanged if we use lags between 2 and 6 months. We also find qualitatively similar results when we adopt a "shift-share" approach that relates migration decisions to prior *changes* in higher-order network structure, holding fixed the direct connections of the migrant.

of omitted variable bias. Our preferred specification includes fixed effects for each individual migrant (to control for individual heterogeneity, for instance that certain people are both more likely to migrate and to have certain types of networks), fixed effects for each possible origin–destination–month combination (to control for factors that are shared by all people facing the same migration decision, such as wage and amenity differentials), and fixed effects for each possible destination network size (such that comparisons are always between places where the migrant has the exact same number of direct contacts, as in Figure 1). Thus, in our preferred specification, the identifying variation comes from within-individual differences in network structure between destinations and over different months in the 5-year window, net the population-average differences that vary by home–destination–month, and net any effects that are common to all people with exactly the same number of friends in the destination. We would observe such variation if, for instance, an individual had been considering a move to a particular destination for several months, but only decided to migrate after his friends in the destination became friends with each other—and if that tightening of his social network exceeded the average tightening of networks in that destination (as might occur around the holidays, for instance).

This analysis helps to establish a new set of stylized facts about the relationship between migration and social networks. Most notably, we show that migrants are more likely to migrate to destinations where their social networks are *interconnected* (*i.e.* where the migrant's friends are friends with each other), but that they are no more likely to migrate to destinations where their networks are *extensive* (*i.e.* where their distance-2 and distance-3 neighbourhoods are larger). In fact, conditional on network size migrants are *less* likely to go to places where their networks are extensive—a result that surprised us initially, given the emphasis prior work has placed on the value of connections to socially distant nodes in a network (*e.g.* Granovetter, 1973). In other words, of the three potential migrants in Figure 1, B is most likely to migrate and C is least likely, with A somewhere in between.

To better understand this "surprising" result, we document considerable heterogeneity in the migration response to social network structure. In particular, we find that the negative effect of extensive networks is strongest when a migrant's direct contacts have a large number of "strong ties" in the destination (where we define a strong tie as one with five or more calls per month, which is equivalent to the 90th percentile of call frequency); when a migrant's destination contacts have many weak ties (*i.e.* ties that are not strong), migration is not deterred. Such evidence suggests that there may be rivalry in information sharing in networks, which leads migrants to value connections to people for whom there is less competition for attention (as in Dunbar, 1998; Banerjee *et al.*, 2012). We also find that while the *average* migrant is not drawn to locations where her friends have more friends (as in G_3), such structure does attract several less common types of migrants. In particular, repeat migrants (who have previously migrated from their home to the destination) and long-term migrants—both of whom are presumably better informed about the structure of the destination network—are more likely to migrate to locations where their networks are more extensive.

To summarize, this paper makes two main contributions. First, it provides a new empirical perspective on the determinants of migration in developing countries (cf. Lucas, 2015). In this literature, many scholars have noted the important role that social networks play in facilitating migration. Early examples in the economics literature include Rees (1966) and Greenwood (1969); a large number of subsequent studies document the empirical relationship between network size and migration rates.[5] More recently, Munshi and Rosenzweig (2016) document

5. Examples include Montgomery (1991), Borjas *et al.* (1992), Munshi (2003), McKenzie and Rapoport (2010), Dolfin and Genicot (2010); Beaman (2012), Patel and Vella (2012), Bertoli *et al.* (2013), and Bertoli and Ruyssen (2018). Two recent papers use phone data to link spatial mobility and social networks. Büchel *et al.* (2020) use data from a Swiss cellphone operator to link migration decisions to phone calls, and document a similar relationship between

that the fear of losing social network ties may prevent profitable migration, while Morten (2019) shows that the act of migration can change social relationships and risk sharing. Kinnan (2019) theorizes about the two-way interconnections: migration of one individual can make other network members better off if that individual has a new source of income, but others may be worse off if the act of migration improves the outside opportunity for that person or makes it easier to hide income. This paper builds on this line of work by exploiting a new source of data to establish a more nuanced set of stylized facts about networks and migration—highlighting, in particular, the value migrants place on interconnected networks, and substantial heterogeneity in how different types of migrants value networks.

Second, through the study of migration, we shed light on the more general question of how social networks provide social capital to individuals embedded in those networks (cf. Jackson, 2010; Banerjee *et al.*, 2013, 2019).[6] We contrast interconnected and extensive networks, just as network theory distinguishes between networks that provide cooperation capital and networks that provide information capital (Jackson, 2020). In that literature, cooperation capital is usually motivated by repeated game models of network interaction, where interconnected networks (*e.g.* cliques) best support social reinforcement and sanctioning.[7] Information capital, which reflects the network's ability to efficiently transmit information, is associated with extensive subnetworks (*e.g.* stars and trees) where an individual is linked to many others via short network paths.[8] We show that, at least to migrants, topologies associated with cooperation capital matter most.

2. DATA AND MEASUREMENT

To study the empirical relationship between networks and migration, we exploit a novel source of data that contains detailed information on the migration histories and evolving social networks of roughly 1 million individuals in Rwanda. These data, obtained from Rwanda's near-monopoly telecommunications company, contain the call detail records (CDR) for all mobile phone activity that occurred in Rwanda from January 2005 until June 2009. For each mobile phone call that occurs, the CDR contain a log of the (anonymized) phone numbers of the two parties involved in the call, a timestamp for when the call was placed, and the identifiers for the cell phone towers through which the call was routed, which in turn indicates the approximate geolocation of each party at the time of the call. In total, we observe roughly 1 billion mobile phone calls between roughly 1 million unique subscribers (Table 1).

By combining information on subscribers' locations (based on the cell towers they use) and social network structure (based on the people they speak to), we are able to study the relationship between migration and social networks. To provide intuition, the network of a single migrant, in the month before migration, is shown in Figure 2. This particular migrant (the green dot) had twenty unique contacts in the month prior to migration, seven of whom were in his home district

network size and migration as the one we note in Section 4.1. Barwick *et al.* (2019) show that migrant flows in a Chinese city correlate with call volume between regions, and link this information flow to improved labour market outcomes. Both papers focus primarily on how on network size relates to migration, whereas our focus is on the role of higher-order network structure, conditional on network size.

6. There is a large literature on social capital that studies how social structure fosters trust and cooperation in a society. In particular, the importance of social pressures on fostering cooperation has deep roots in the sociology literature (cf. seminal work by Simmel (1950) and Coleman (1988), among many others).

7. Jackson *et al.* (2012) and Ali and Miller (2016) provide recent examples. See also Ligon and Schechter (2012), Jackson *et al.* (2012), Ambrus *et al.* (2015), and Chandrasekhar *et al.* (2018).

8. Early models include Kermack and McKendrick (1927) and Jackson and Wolinsky (1996); more recent examples include Calvó-Armengol and Jackson (2004), Jackson and Yariv (2010), and Banerjee *et al.* (2013).

TABLE 1
Summary statistics of mobile phone metadata

	(1) In a single month (January 2008)	(2) Over 2 years (July 2006–June 2008)
Number of unique individuals	432,642	793,791
Number of CDR transactions	50,738,365	868,709,410
Number of migrations	21,182	263,208
Number of rural-to-rural migrations	11,316	130,009
Number of rural-to-urban migrations	4,908	66,935
Number of urban-to-rural migrations	4,958	66,264

Notes: Migration statistics calculated from Rwandan mobile phone data. Column (1) is based on data from a single month; column (2) includes 2 years of data, potentially counting each individual more than once. A "migration" in this table is defined as occurring when an individual remains in one district for 2 consecutive months and then remains in a different district for the next 2 consecutive months. We denote the three districts in the capital of Kigali as urban; the remaining districts are considered rural. Supplementary Table A1 provides additional summary statistics for different types of migration events.

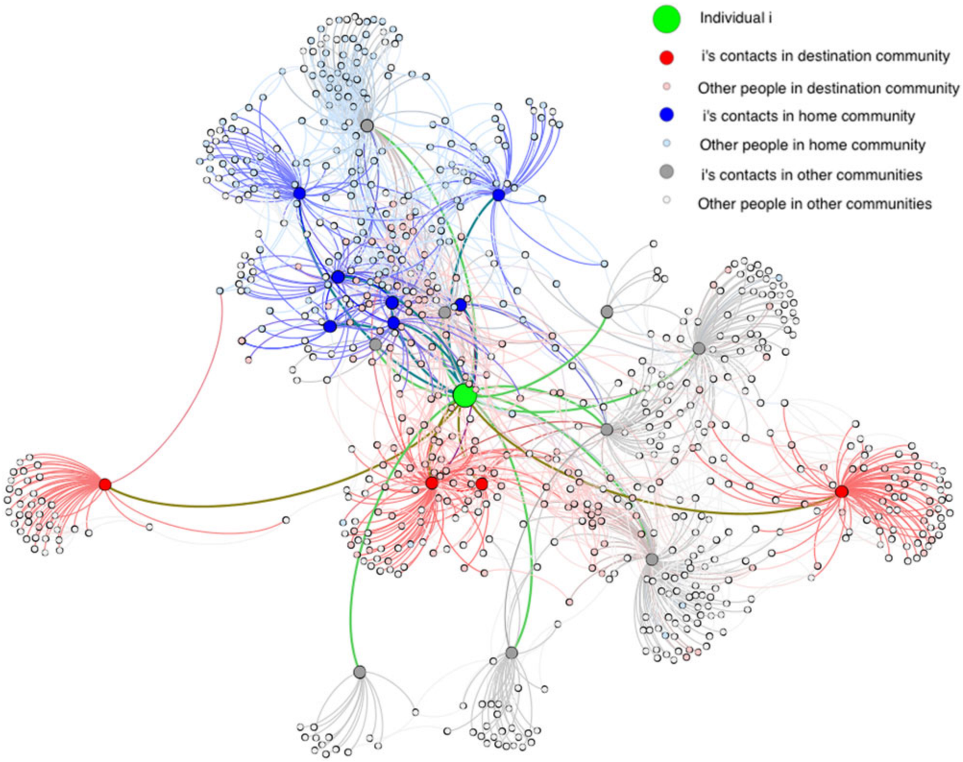

FIGURE 2
The social network of a single migrant

Notes: Diagram shows the social network, as inferred from phone records, of a single migrant *i*. Nodes represent individuals; edges indicate that two individuals communicated in the month prior to *i*'s migration. Direct contacts of *i* are shown in blue (for people *i*'s home district), red (for people in *i*'s destination district), and solid grey (for people in other districts). Small hollow circles indicate *i*'s "friends of friends," that is, people who are not direct contacts of *i*, but who are direct contacts of *i*'s contacts. All individuals within two hops of *i* are shown. Nodes are spaced using the force-directed algorithm described in Hu (2005).

(blue dots), four of whom were in the destination district (red dots), and the remainder were in other districts (grey dots). Friends of friends are depicted as hollow grey circles.[9]

This section describes how we use these data to observe the structure of each individual's social network over time (Section 2.1) and to extract each individual's complete migration history (Section 2.2). Section 2.3 discusses limitations of these data. In our empirical analysis, we remove personally identifying information (including phone numbers) from the CDR we received from the mobile operator. In addition, to focus our analysis on individuals rather than businesses, and to remove the potential impact of spammers and call centres, we remove all transactions involving numbers with more than 200 unique contacts in a single month (this represents the 95th percentile). In later robustness checks, we confirm that these thresholds for outlier removal do not affect our results.

2.1. *Modelling social network structure with mobile phone data*

Our central goal is to understand how the structure of an individual's social network in different geographic locations affects their likelihood of migrating to and from those locations. The social network we observe is that of mobile phone communications. Specifically, we use the set of calls occurring within a specific time frame (typically a month) to define the (undirected) social network G at that time. Formally, let the network G be a matrix with $G_{ij} = G_{ji} = 1$ if i and j are observed to talk on the phone within a fixed time window and $G_{ij} = G_{ji} = 0$ otherwise (this includes $G_{ii} = 0$). A *path* between i and j is an ordered sequence of distinct agents $(i i_1 i_2 \ldots i_h j)$ such that any two adjacent agents are connected in the network. The *distance* between i and j, denoted as $d(i, j)$, is the length of the shortest path between i and j.[10]

Since network structure can be quite complex (as in Figure 2), we focus on two archetypal ways that the *topology* of the social network influences the decision to migrate: through *information capital* and *cooperation capital*. Following Jackson (2020), we think of information capital as the potential for the social network to provide access to novel information—about jobs, new opportunities, and the like. By contrast, we consider the cooperation capital as the network's ability to facilitate interactions that benefit from cooperation and community enforcement, such as risk sharing and social insurance (similar to the notion of *favour capital* in Jackson, 2020).

2.1.1. Information capital. We construct a proxy measure of information capital for agent i in network G by measuring the size of the agent's second-degree neighbourhood (or unique *friends of friends*, not counting direct contacts of the agent):

$$D_i^2(G) = \{j : d(i, j) = 2\}. \tag{1}$$

More generally, agent i's kth-degree neighbourhood is $D_i^k(G) = \{j : d(i, j) = k\}$.

In Supplementary Appendix A1, we develop a model that provides microeconomic foundations for this particular measure of information capital. Broadly speaking, it is intended to capture the *extensiveness* of the agent's network, that is, the extent to which one person is linked to many others via short network paths (cf. Granovetter, 1973). This measure relates closely to Jackson and Wolinsky's (1996) notion of decay centrality and Banerjee *et al.*'s (2013) measure of diffusion centrality. With both decay and diffusion centrality, information capital

9. Throughout, we use the term "friend" loosely, to refer to the contacts we observe in the mobile phone network. These contacts may be friends, family, business relations, or something else.

10. If there is no path between i and j, then $d(i, j) = \infty$.

increases with more friends, friends of friends, and so on.[11] As the number of friends is an agent's endogenous choice which we will examine separately, the number of unique friends of friends—the measure defined in equation (1)—is the key proxy to capture the individual's access to information in the outer network.

2.1.2. Cooperation capital. Our main proxy for an individual's cooperation capital is *network support*, that is, the probability that an agent i's friend j has one or more friends in common with i. Formally, agent i's support in network G is 0 if i does not have any friend in G, otherwise

$$\text{Support}_i(G) \equiv \frac{\#\{j : G_{ij} = 1 \ \& \ (G^2)_{ij} \geq 1\}}{\#\{j : G_{ij} = 1\}}. \tag{2}$$

This proxy is designed to measure the *interconnectedness* of the agent's network, and relates closely to the notion of *favour capital* in Jackson (2020), defined as the network's "ability to exchange favours and transact with others through network position and repeated interaction and reciprocation" (p. 315). Importantly, cooperation capital is facilitated by different network topologies than information capital: Supplementary Appendix A1 shows that group enforcement is strong and cooperation is efficient when local sub-networks have high levels of support.[12] Network support is also correlated with *network clustering* (the probability that two friends are connected to each other), a metric we use in later tests of robustness.

2.1.3. Summary. Our empirical analysis distills the complex structure of social networks into two stylized network statistics: *Unique friends of friends*, a proxy for information capital, defined by equation (1); and *% Friends with common support*, a proxy for cooperation capital, defined by equation (2). We will also show results pertaining to i's *degree centrality*, $|D_i^1(G)|$, which simply counts the number of unique individuals with whom each person communicates. Ancillary results will separately account for the *strength* of a social tie, which we measure as the number of (undirected) calls between two individuals; when we compare strong and weak ties, we consider "strong" ties to be those ties in the 90th percentile of the tie strength distribution (equivalent to five or more calls per month).

For most of the analysis that follows, we partition the full social network of Rwandan mobile subscribers (containing approximately 800,000 individuals) into twenty-seven location-specific sub-networks, each of which is defined by the administrative districts of Rwanda.[13] Thus, we calculate equations (1) and (2) separately for each sub-network G_d, which only has entries for individuals who reside in d. This simplifying assumption dramatically expedites our computational analysis, but assumes that agent i cannot derive social capital from a given district d via

11. Jackson (2020) describes information capital as "ability to acquire valuable information and/or spread it to other people through social connections" (p. 4). Decay centrality (Jackson and Wolinsky, 1996) assumes each agent receives a value $q < 1$ (the probability of information transmission) from each direct friend, a discounted value of q^2 from each friend of friend, and so on. Diffusion centrality (Banerjee *et al.*, 2013) further accounts for the fact that multiple paths could increase the chance that information makes it from one individual to other.

12. See also the references in footnote 7. In particular, Ali and Miller (2016) model a dynamic game of repeated cooperation and find that a clique network (a completely connected network) generates more cooperation and higher average cooperation capital than any other networks; Jackson *et al.* (2012) model a game of repeated favour exchanges and highlight the importance of *supported* relationships, where a link is supported if the two agents of the link share at least one common friend.

13. Our analysis groups the three smaller and contiguous districts that comprise the capital of Kigali into one "district".

people residing outside d. We therefore include analysis that shows how our main results are affected by relaxing this assumption (see Section 4.3).

2.2. *Modelling migration*

2.2.1. Internal migration in Rwanda. Internal migration is a prominent feature of most developing countries. According to the United Nations Population Division (2013), there are an estimated 762 million internal migrants in the world. Yet, survey-based data on internal migration are notoriously unreliable, particularly in developing countries where many migrations are temporary (Deshingkar and Grimm, 2005; McKenzie and Sasin, 2007; Carletto *et al.*, 2012; Lucas, 2015).

Our empirical analysis focuses on internal migration in Rwanda, a small agricultural economy in East Africa. Rwanda has high rates of poverty, estimated by the National Institute of Statistics of Rwanda (2012) to be 56.7% in 2005 (the beginning of the period we study). While fewer than 4% of Rwandan residents are born abroad, internal migration in Rwanda is common. According to the National Institute of Statistics of Rwanda (2014), roughly 20% of the resident population has experienced a lifetime migration, with similar migration rates for men and women. As with many predominantly agricultural societies, the most frequent type of internal migration in Rwanda is from one rural location to another Lucas (2015). For instance, the World Bank estimates that between 2005 and 2011, roughly two-thirds of all migrants went to rural destinations; less than 20% of migrants were from rural-to-urban areas (World Bank Group, 2017).

The push and pull factors driving internal migration in Rwanda have varied over the last few decades. The 1994 genocide and surrounding conflict were major drivers of internal migration in the 1990s, but conflict has been far less common since 2000. While the National Institute of Statistics of Rwanda (2014) did not collect data on migration motives, their analysis of patterns of urban and rural migration by gender "suggests that males mainly migrate towards urban areas for employment purposes while women tend to move shorter distances, either for marriage or agricultural purposes" (p. 7). Likewise, a series of reports from the Famine Early Warning System highlights the role that agriculture and construction play in driving labour migration, but also emphasizes the unpredictability of this demand (FEWS NET Rwanda, 2014). A more comprehensive study of internal migration in Rwanda, conducted by the World Bank Group (2017) and based on nationally representative household survey data from 2014 (EICV4), notes other factors common to many African countries: rural-to-urban migrations are driven by urban employment opportunities and rural land shortages, and urban-to-rural migrations are frequently motivated by the high cost of living in the city and a desire for lower density areas where farmland may be available (especially in the Eastern Province).

2.2.2. Measuring migration with mobile phone data. We use mobile phone metadata to provide a detailed quantitative perspective on the migration patterns of mobile phone owners in Rwanda. This is possible because each time a mobile phone call is placed, the operator logs the cell phone towers through which the caller and receiver were connected; typically, these are the towers closest to the subscribers at the time of the call. As can be seen in Figure 3, this allows us to approximately and intermittently locate each subscriber, with a geographic precision of a few hundred metres in urban areas and several kilometres in rural areas.

We use the sequence of mobile phone towers to reconstruct each individual's history of migration. Our approach, described in more detail in Supplementary Appendix A3, builds on prior methodological work to infer migration from mobile phone data (cf. Blumenstock, 2012; Lai *et al.*, 2019; Chi *et al.*, 2020). To summarize, we first calculate the district of residence d

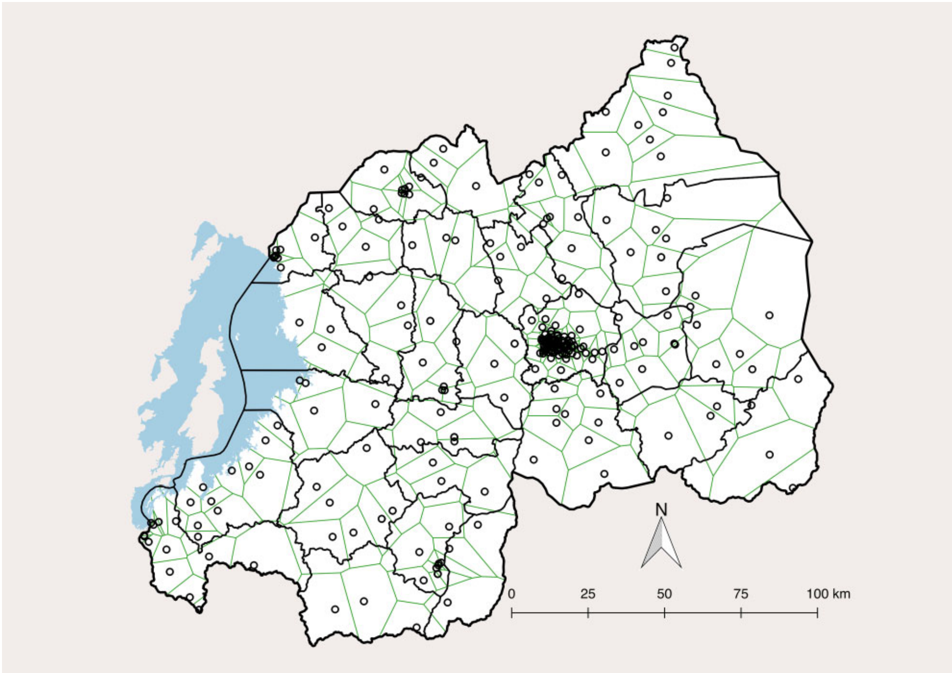

FIGURE 3
Location of all mobile phone towers in Rwanda, circa 2008

Notes: Circles indicate cell tower locations. Dark/black lines represent district borders. Light/green lines show the voronoi polygons roughly divide the country into the coverage region of each tower.

of every individual i in each calendar month t (districts are shown with black lines in Figure 3). We do this using Supplementary Algorithm 1, which determines the district in which i spent the majority of evenings during t.[14] From this sequence of monthly residential locations, we then determine whether or not each individual migrates in each month. Following Blumenstock (2012), we say that a migration occurs in month $t+1$ if three conditions are met: (1) the individual's home location is observed in district d for at least k months prior to (and including) t; (2) the home location d' in $t+1$ is different from d; and (3) the individual's new home location is observed to be district d' for at least k months after (and including) $t+1$. Individuals whose home location is observed to be in d for at least k months both before and after t are considered residents, or stayers. Individuals who do not meet these conditions (such as individuals who do not use their phone for an entire month, or individuals who do not remain in one district for k consecutive months) are treated as "other" (and are excluded from later analysis).

Using this approach, we are able to provide granular detail on patterns of internal migration in Rwanda. Table 1 provides summary statistics on internal migration events observed in the mobile phone data, where a migration is defined as an instance where an individual stays in one district for at least 2 months, then moves to a new district and remains in that new district for at least 2 months (*i.e.* $k=2$). The first column highlights data from a single month; the

14. For each evening hour, we infer i's location as the district in which a majority of their calls occurred. We then infer i's location for the entire evening as the district in which the evening hours were spent. Ties are resolved with a coin toss.

second column aggregates over a 2-year period. Supplementary Table A1 includes summary statistics for alternate definitions of migration—including different values of k and specific types of migration that are prominent in the literature on internal migration in developing countries (cf. Todaro, 1980; Lucas, 1997, 2015). Broadly, we observe a large number of repeat and circular migrants, with a majority of migrants traveling long distances. We also note that, comparing the rows of Supplementary Table A1, the implied migration rate decreases as the minimum stay length k is increased. Such comparisons would be difficult with traditional survey data, which typically capture a single definition of migration. In later analysis, we use a minimum stay length of $k = 2$ as our base definition of migration, as the implied migration rate roughly matches official statistics on internal migration provided by the Rwandan government (National Institute of Statistics of Rwanda, 2014).

2.3. *Data limitations and validation*

While mobile phone data provide uniquely granular insight into the social networks and migration decisions of a large population, they also have several important limitations.

2.3.1. Non-representative population. During the period under study (early 2005–early 2009), mobile phone penetration rose from roughly 5–22% (estimates based on the number of subscribers who appear in our dataset). During this time, mobile phone subscribers in Rwanda were not representative of the larger Rwandan population; survey evidence suggests they were significantly wealthier, older, better educated, and are more likely to be male (Blumenstock and Eagle, 2012). While this non-representativeness limits the external validity of our analysis, survey evidence suggests that the population of phone owners and the population of migrants have similar demographic characteristics.[15] More importantly, our empirical specifications are designed to limit the scope for patterns of phone ownership, including trends in mobile penetration over time, to bias our results—see the discussion of omitted variable bias and shift-share analysis in Section 3.3.

2.3.2. Phones are not people. The unique identifiers we observe are for mobile phone numbers, not individuals. As noted above, we attempt to limit the extent to which firms and organizations influence our analysis by removing phones with unusually high activity (as well as any traffic associated with those phones). Still, when multiple people share the same phone number, we may overestimate the size of an individual's network. It is also possible that a single individual might use multiple phone numbers, but we believe this was less common since there was only one dominant phone operator at this time. In principle, our data make it possible to uniquely identify devices and SIM cards, in addition to phone numbers. Of these, we believe that phone numbers (which is portable across devices and SIM cards) most closely correspond to unique individuals.

2.3.3. Construct validity. The social network we observe is the network of mobile phone relations, which is a subset of all true social relations in Rwanda. This subset is non-random: it is biased towards certain demographic groups; it systematically understates certain types of

15. For instance, the age distribution of migrants estimated from 2012 government census data (National Institute of Statistics of Rwanda, 2014, Figure 11) is similar to the age distribution of a representative survey of mobile phone owners in 2009 (Blumenstock and Eagle, 2010, Figure 2).

relationships (such as those that are primarily face-to-face); and may overstate other more transient or functional relationships (such as with a shopkeeper). We address some of these concerns through robustness tests that vary the definition of "social tie," for instance by only considering edges with several observed communication events (see Section 4.4). Other concerns are ameliorated by the fact that much of our analysis focuses on long-distance relationships, and during this period in Rwanda the mobile phone was the primary means of communicating over distance.

Related, we measure migration based on the movement of phones, rather than with traditional survey-based instruments. Prior work suggests that patterns of migration inferred from mobile phone data broadly match inferences drawn from other sources—this includes work in Rwanda using the same dataset as in this paper (Blumenstock, 2012; Williams *et al.*, 2013), neighbouring countries in East Africa (Wesolowski, Buckee, *et al.*, 2013; Wesolowski, Eagle, *et al.*, 2013; Pindolia *et al.*, 2014), as well as other low-income (Bengtsson *et al.*, 2011; Lu *et al.*, 2016; Lai *et al.*, 2019) and wealthy nations (Lenormand *et al.*, 2014). In our context, the aggregate patterns of population flows that we calculate from the mobile phone data between 2005 and 2009 are broadly similar to those reported in the 2012 Rwandan census, but there are discrepancies between the two measurements. For instance, Supplementary Figure A1 compares estimates of internal migration from the phone data (red bars) to those from the census (blue bars), as reported by National Institute of Statistics of Rwanda (2014, p. 29). These inconsistencies could be due to the non-representativeness of phone owners, to differences in how the two instruments define migration,[16] or to the fact that mobile phone data is a more sensitive instrument for detecting human mobility than the typical census questionnaire.

3. IDENTIFICATION AND ESTIMATION

The focus of this paper is on understanding how social networks influence the decision to migrate. While a host of other factors also influence that decision—from wage and amenity differentials to physical distance and associated migration costs—we study how, holding all such factors fixed, variation in social network structure systematically correlates with migration decisions. In the stylized example of Figure 1, we ask whether a person with network G_1 is more likely to migrate than someone with network G_2, whose network is marginally more interconnected and would be expected to provide marginally more cooperation capital. We similarly compare the migration decisions of such individuals to individuals with network G_3, which is slightly more extensive and would be expected to provide slightly more information capital. In practice, of course, the actual network structures are much more complex (as in Figure 2). We therefore use statistical models to estimate the effect of marginal changes in complex network structure on subsequent migration decisions.

The central difficulty in identifying the causal effect of social networks on migration is that the social networks we observe are not exogenous: people migrate to places where their networks have certain characteristics, but this does not imply that the network caused them to go there. Here, we describe our estimation strategy, and the identifying assumptions required to interpret our estimates as causal.

16. Our algorithm defines a migrant as someone who remains in one district for 2 or more months and then moves to another district for 2 or more months. The Rwandan census does not capture this type of short-term migration; we instead show the census estimates of internal recent migrants, which are defined as "a person who moved to his/her current district of residence 5 years or less prior to the census".

3.1. *Simultaneity*

An obstacle to understanding the causal effect of networks on migration is that migration decisions may also shape networks. This would be expected if, for instance, migrants strategically formed links to destination communities in anticipation of migration, or simply made a large number of phone calls to their destination before migrating.

We address this concern in three principal ways. First, we study the effect of lagged network characteristics on the current decisions of migrants. Specifically, we relate the migration decision made by individual i in month t to the structure of i's social network s months prior. As a concrete example, when $t =$ August 2008 and $s = 2$, we relate the August 2008 migration decision to the structure of the individual's social network in June 2008. Our main specifications use $s = 2$, but we later show that our results are unchanged when with longer lags. Second, rather than focus on the *number* of direct contacts a migrant has at home and in the destination, we focus on the *connections* of those contacts, holding the number of contacts fixed (as in Figure 1). This is because it seems easier for a migrant to directly control the number of contacts they have in the destination and at home than it is for them to alter the higher-order structure of their social network. Third, in tests described in Section 4.4, we adopt a "shift-share" specification that relates migration decisions to *changes* in an individual's higher-order network (for instance, between $t - 12$ and $t - 2$), holding lower-order network structure fixed, in order to further limit the extent to which the individual could endogenously shape their network.

These techniques reduce, but do not eliminate, the potential for simultaneity. In particular, a migrant might plan her migration many months in advance of migration, and in that process could change her higher-order network structure—for instance by asking a friend to make new friends on her behalf, or by encouraging two friends to talk to each other. To gauge the extent to which this might bias our results, we run several empirical tests, and find little evidence of such anticipatory behaviour. For instance, Figure 4(a) shows, for a random sample of migrants, how the geographic distribution of migrants' social networks changes over time. Prior to migration, roughly 40% of the average migrant's contacts are in the origin and 25% are in the destination; 3 months after migration, these proportions have switched, reflecting how the migrant has adapted to her new community. Notably, however, migrants do not appear to strategically form contacts in the destination immediately prior to migrating; if anything, migrants shift their focus to the people in the community they are leaving—and any deviations from the long-term trend do not appear until the month of migration. These compositional changes do not mask a systematic increase in the *number* of contacts in the destination, or the number of total calls to the destination: Supplementary Figure A2 indicates that the total number of contacts increases over time, but there is no sudden spike in the destination district in the months before migration; Supplementary Figure A3 shows analogous results for total call volume. As a sort of "placebo" test, Figure 4(b) shows how networks evolve over time for non-migrants, where we draw a sample non-migrants that matches the temporal distribution of migrants from Figure 4(a).[17] While non-migrants have different network structure than migrants (in particular, non-migrants have a higher fraction of contacts in their home district than migrants do), there are no sudden changes in the composition of network structure of non-migrants. With non-migrants, as was the case with migrants, we observe a slow long-term trend towards a larger share of communication being within the home district.

What matters most to our identification strategy is that we similarly find no evidence that migrants are systematically altering the *higher-order* structure of their social networks in the

17. Specifically, for each migrant who appears in Figure 4(a) and is observed to migrate in month t, we randomly sample a non-migrant from the same month t to include in Figure 4(b).

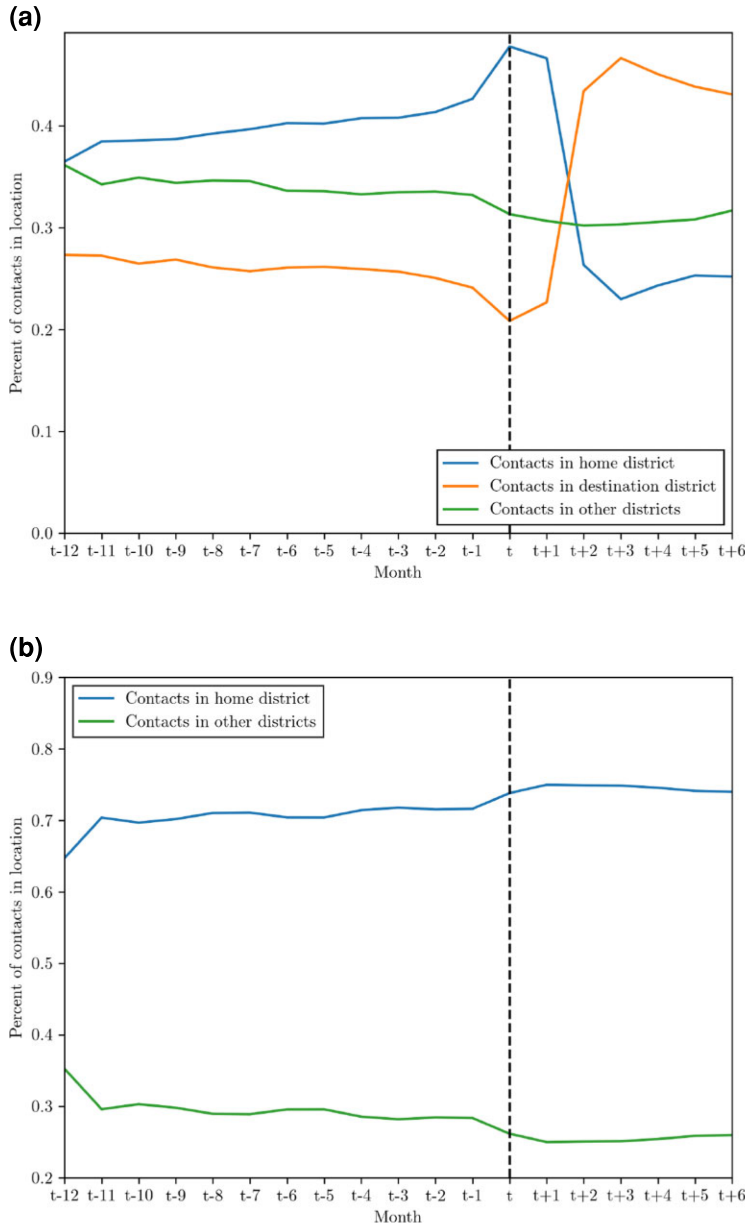

FIGURE 4

Changes in network structure over time. (a) Migrants. (b) Non-migrants

Notes: Figures show how the network connections of (a) migrants and (b) non-migrants evolves over time. For (a), we draw a random sample of 10,000 migrants, and plot the average percentage of contacts those individuals have in the home, destination, and other districts, in each of the 12 months before and 6 months after migration. The dashed vertical line indicates the date of migration. For (b), we draw a random sample of 10,000 non-migrants by selecting, for each migrant who is sampled to appear in (a) and observed to migrate in month t, a non-migrant from the same reference month t.

months prior to migration. In particular, Supplementary Figure A4 indicates that while the number of migrants' friends of friends slightly increases over time, there is no noticeable shift in the months prior to migration. Supplementary Figure A5 shows similar results for the level of common support in the network.[18]

3.2. *Omitted variables*

The second main threat to identification is the fact that network structure may be a proxy for other characteristics of the individual (*e.g.* wealth, ethnicity) and location (*e.g.* population density, wages) that also influence migration. Our main strategy for dealing with such omitted variables is to include an extremely restrictive set of fixed effects that control for many of the most concerning sources of endogeneity. This strategy is possible because of the sheer volume of data at our disposal, which allow us to condition on factors that would be impossible in regressions using traditional survey-based migration data.

Our preferred specification includes fixed effects for each individual (roughly 800,000 fixed effects), for each origin–destination–month tuple (roughly 18,000 fixed effects), and for the number of direct contacts in the destination (roughly 100 fixed effects). The individual fixed effects absorb all time-invariant individual heterogeneity (such as wealth, gender, ethnicity, personality type, family structure, and so forth), and addresses the fact that some people are inherently more likely to migrate than others (and have inherently different social networks). The origin–destination–month fixed effects control for any factor that similarly affects all individuals considering the same origin–destination migration in the same month. This includes factors such as physical distance, the cost of a bus ticket, location-specific amenities that all migrants value equally, average wage differentials, and many of the other key determinants of migration documented in the literature (including the usual "gravity" effects in a standard trade or migration model).[19] Finally, we include fixed effects for the number of first-degree contacts in the destination in order to isolate the effect of differences in higher-order network structure on migration.

3.3. *Identification*

To summarize, the identifying variation in our main specification is (1) within-individual over time and (2) within-individual over potential destinations—in both instances, after controlling for any factors that are shared by all people considering the same destination in the same month, and for any effects that are common to all people with the same number of direct contacts in the destination. An example of (1) could occur if, for instance, an individual had been considering a move to a specific destination for several months, but only decided to migrate after his friends in the destination became friends with each other (the G_2 vs. G_1 comparison of Figure 1)—and if the increased interconnectedness exceeded the average increase of networks in that destination (as might occur around the holidays, for instance). An example of (2) could occur if, in a given

18. In Section 4.4, we further show that there are no sudden changes in higher-order structure even after "freezing" the migrant's contacts at $t-12$.

19. For instance, we know that rates of migration are higher to urban centres, and that social networks in urban centres look different from rural networks. Including a destination fixed effect removes all such variation from the identifying variation used to estimate the effect of networks on migration. The origin–destination–month fixed effects remove destination-specific variation, as well as more complex confounding factors that vary by destination and origin and time, such as the possibility that the seasonal wage differential between two districts correlates with (lagged) fluctuations in social network structure.

month, a single migrant were choosing between two destination districts, had the same number of contacts in each district, and then decided to migrate to the district where his contacts were more interconnected. Prima facie, it may seem unlikely that such small differences would shape the decision to migrate, but our data allow us to ascertain whether, across millions of individual migration decisions, such a general tendency exists.

The fixed effects we include significantly reduce the scope for omitted variables to bias our estimates of the effect of network structure on migration, but they do not eliminate such bias entirely. If, for instance, origin–destination wage differentials are individual-specific, our fixed effects will not help. This might occur if carpenters' networks in a particular district grew more interconnected over time (relative to carpenter network growth in other districts) than farmers' networks in that district (again relative to farmers' networks in other locations), and if migration rates of carpenters to that district are higher for reasons unrelated to network structure. Likewise, our use of lagged network structure reduces, but does not eliminate, the likelihood that a migrant would first decide to migrate and then modify his network accordingly.

We revisit these concerns, and other possible threats to identification, in Section 4.4, after introducing the estimation strategy and presenting the main results. In Section 4.4, we precisely state the identifying assumption, discuss the most likely threats to identification, and perform a number of tests to assess the plausibility of this identification strategy.

3.4. *Estimation*

Formally, for an individual i considering a move from home district h to destination district d in month t, we wish to estimate the effect of a vector of (s-lagged) network characteristics $Z_{id(t-s)}$ on the migration decision. This is a discrete choice setting in which i faces twenty-seven mutually exclusive choices in month t, one for each district d in Rwanda (including the home district h). We assume the indirect social capital i would receive from being in d is a function of individual characteristics (μ_i), fixed characteristics of d in the month the choice is being made (π_{dt} and ν_{dt} for destination and home districts, respectively), and a vector of choice-specific attributes that may also differ across individuals ($Z_{id(t-s)}$):

$$U_{idt} = \mathbb{1}(d \neq h)\left[\beta_d' Z_{id(t-s)} + \pi_{dt}\right] + \mathbb{1}(d = h)\left[\gamma + \beta_h' Z_{id(t-s)} + \nu_{dt}\right] + \mu_i + \epsilon_{idt}. \quad (3)$$

Our focus is on the influence of i's network $Z_{id(t-s)}$ (measured with s lags), which the above formulation allows to differ for home networks ($d = h$) and destination networks ($d \neq h$). The vectors β_d and β_h contain the coefficients of interest, which indicate the average effect of destination and home network properties, respectively, on the probability of migration. The parameter γ captures the average tendency for individuals to not migrate.

Assuming that ϵ_{idt} is drawn from an extreme value distribution, i will choose d at time t with probability:

$$P(M_{idt} = 1) = \frac{\exp(\tilde{U}_{idt})}{\Sigma_{d'} \exp(\tilde{U}_{id't})}$$

which can be estimated with a conditional logit model (using $\tilde{U}$ to denote U without the disturbance term ϵ).[20] We omit μ_i from $\tilde{U}$ because it does not vary across the set of choices faced by i in month t.

20. We use the approach described by Eaton *et al.* (2012) and Sotelo (2019) to estimate the multinomial model by Poisson pseudo-maximum likelihood—see also Correia *et al.* (2019).

Since much of our focus will be on understanding how the shape of an individual's *higher-order* network structure relates to their decision to migrate, many specifications will additionally control flexibly for the size of i's first-degree network:

$$U_{idt} = \mathbb{1}(d \neq h)\Big[\beta_d' Z_{id(t-s)} + \pi_{dt} + \sum_k \eta_k \mathbb{1}(D_{id(t-s)} = k)\Big] + \mathbb{1}(d = h)\Big[\gamma + \beta_h' Z_{id(t-s)} + \nu_{dt} + \sum_k \zeta_k \mathbb{1}(D_{id(t-s)} = k)\Big] + \mu_i + \epsilon_{idt}. \quad (4)$$

In the above specification, the vectors of fixed effects η_k and ζ_k allow for migration probabilities to differ for people with different numbers of unique contacts k both at home and in the destination.

When estimating equations (3) and (4), individuals are only considered in months where they can be classified as a migrant or a non-migrant in that month. When an individual is classified as "other" (see Section 2.2), those observations are excluded from the regression. Except as noted, specifications use cluster robust standard errors, clustered by individual. Alternative treatments of the standard errors are discussed in Section 4.4.

4. RESULTS

4.1. *The effect of network size in the destination and at home*

Table 2 summarizes the main results from estimating the multinomial logit model (3) of the migration decision on lagged network structure. We find that on average, each additional contact in the destination is associated with a 0.316% increase in the likelihood of migration to that destination (column (2)), and each contact at home is associated with a 0.081% decrease in the likelihood of migration. As discussed in Section 3.3, these coefficients are identified by changes within the individual's network over time, and across destinations in a single period. Comparing the coefficients on Destination degree and Home degree in the first two columns, we can compare the "push" and "pull" forces of networks on migration (cf. Hare, 1999): the effect of adding one additional contact in the destination is roughly 4–6 times as important as an additional contact at home.

The coefficients in the first row of Table 2 validate a central thesis of prior research on networks and migration, which is that individuals are more likely to migrate to places where they have more connections. But the richness of our data allow us to do much more than simply look at these average effects. For instance, Figure 5(a) shows how the average migration rate varies by degree centrality at destination (*i.e.* the number of unique contacts of the individual). We observe that, for instance, roughly 4% of individuals with ten contacts in a potential destination d in month $t-2$ migrate to that location at t. More broadly, we observe that the relationship between migration and network size is positive, monotonic, and approximately linear with slope of unity.

Just as migrants appear drawn to destinations where they have a large number of contacts, migrants are less likely to leave origins where they have a large number of contacts. Figure 5(b) shows the monotonically decreasing relationship between migration rates and the individual's degree centrality at home, where the probability of leaving home decreases in proportion to the size of the home network.

TABLE 2

Effects of home and destination network structure on migration

	(1)	(2)	(3)
Destination degree (network size)	260.24***	315.86***	
	(2.48)	(2.53)	
Destination % friends with support	2586.98***	2270.11***	182.84***
	(11.50)	(12.03)	(11.75)
Destination friends of friends	−2.16***	−5.03***	−0.14***
	(0.08)	(0.09)	(0.06)
Home degree (network size)	41.67***	81.34***	
	(1.15)	(1.27)	
Home % friends with support	845.24***	815.48***	105.62***
	(17.17)	(17.38)	(17.89)
Home friends of friends	2.35***	0.52***	1.38***
	(0.04)	(0.05)	(0.04)
Home district	8229.08***		
	(9.54)		
Observations	184,637,637	184,637,637	184,637,637
Pseudo-R^2	0.68	0.68	0.68
Degree fixed effects	No	No	Yes
Destination × Month fixed effects	No	Yes	Yes
Individual × Month fixed effects	Yes	Yes	Yes

Notes: Each observation corresponds to an individual–month–district tuple. Each column indicates a separate regression of a binary variable indicating 1 if an individual i chose to live in district d in month t. Results are estimated using a conditional logit model, using social network characteristics of the location calculated in month $t-2$. "Home district" is a binary variable indicating whether the destination choice in t is i's home in $t-1$. Degree fixed effects, as well as Destination × Month fixed effects, are interacted with the Home district fixed effect. See discussion in Section 3.4. Coefficients and standard errors are multiplied by 1,000 to make the tables more readable. Standard errors are clustered by individual. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

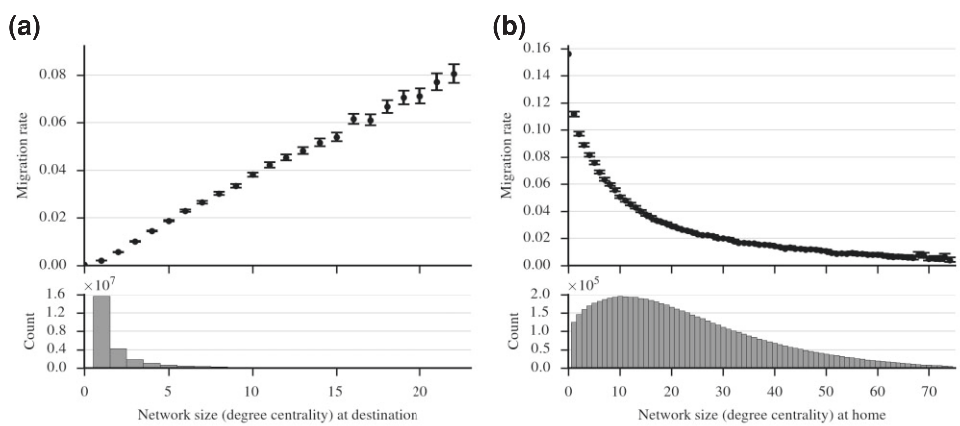

FIGURE 5

Migration and degree centrality (number of unique contacts in network). (a) Degree centrality at destination. (b) Degree centrality at home

Notes: Lower histograms indicate the unconditional degree distribution, that is, the number of individual–month observations for each degree centrality (*i.e.* the number of unique contacts) in the (a) destination network and (b) home network. The upper figures show, at each level of degree centrality (in month $t-2$), the average migration rate (in month t). Error bars indicate 95% confidence intervals, using the Wilson Score interval for binomial proportions.

4.2. *Higher-order network structure*

We next examine the role of *higher-order* network structure—that is, the connections of the individual's contacts—in migration decisions. This analysis uses specification (4), which includes fixed effects for each possible network size, so that identification now comes from changes (over time and across destinations) in the *inter*-connections of the migrant's network (*i.e.* the connections of i's connections), holding the number of direct contacts fixed. Results in the second and third rows of Table 2 highlight the two main results that we explore in greater detail below: migrants are more likely to go to places where their destination networks are more interconnected (row 2); but they are in fact *less* likely to migrate to destinations where their contacts have a large number of contacts (row 3).[21]

4.2.1. Network "interconnectedness". The results in Table 2 indicate that, on average, migrants are more likely to go do destinations that are more interconnected. In other words, networks like G_2 in Figure 1 are more attractive than networks like G_1.[22] Our data allow us to disaggregate this effect, and unpack how migration rates vary at different levels of network interconnectedness. In particular, the left panels (a and c) of Figure 6 show how the average migration rates varies with network support, the measure of interconnectedness defined by equation (2). The lack of a clear trend in the left panels is difficult to interpret in part because network support can vary systematically with network size. For this reason, the right panels (b and d) of Figure 6 show how migration varies with network support, *holding network size fixed.*

Specifically, the right panels of Figure 6 are produced by plotting the β_{kd} and β_{kh} coefficients from:

$$\begin{aligned} U_{idt} = \mathbb{1}(d \neq h)\left[\sum_k \mathbb{1}(D_{id(t-s)} = k)[\eta_k + \beta'_{kd} Z_{id(t-s)}] + \pi_{dt}\right] \\ + \mathbb{1}(d = h)\left[\gamma + \sum_k \mathbb{1}(D_{id(t-s)} = k)[\zeta_k + \beta'_{kh} Z_{id(t-s)}] + \nu_{dt}\right] + \mu_i + \epsilon_{idt}. \quad (5) \end{aligned}$$

This specification is directly analogous to specification (4), which was used to estimate column (3) of Table 2. The key difference is that where specification (4) provided a single estimate of the average effect of network support on migration, specification (5) estimates the effect of network support $Z_{id(t-s)}$ separately for each unique value of network size k.

Panels (b) and (d) of Figure 6 reinforce the prior finding that people are systematically drawn to places where their networks are more interconnected: most coefficients in Figure 6(b) and (d) are (weakly) positive, indicating that migrants with a wide variety of network sizes are drawn to places where those networks are more interconnected. The figure also adds a level of nuance that would not be possible with traditional survey-based data. For instance, the fact that the β_{kd} coefficients in Figure 6(b) are generally increasing indicates that as the potential migrant

21. Similar results are obtained when focusing on the choice of destination among individuals who have already decided to migrate: Supplementary Table A2 presents multinomial logit results estimated just on the subset of individual–month where the individual is observed to migrate. The results across columns are qualitatively unchanged from the main results in Table 2.

22. It is worth noting that in other settings, more network interconnections are not necessarily attractive. For instance, Ugander *et al.* (2012) show that people are less likely to sign up for Facebook when their pre-existing Facebook friend network is more interconnected.

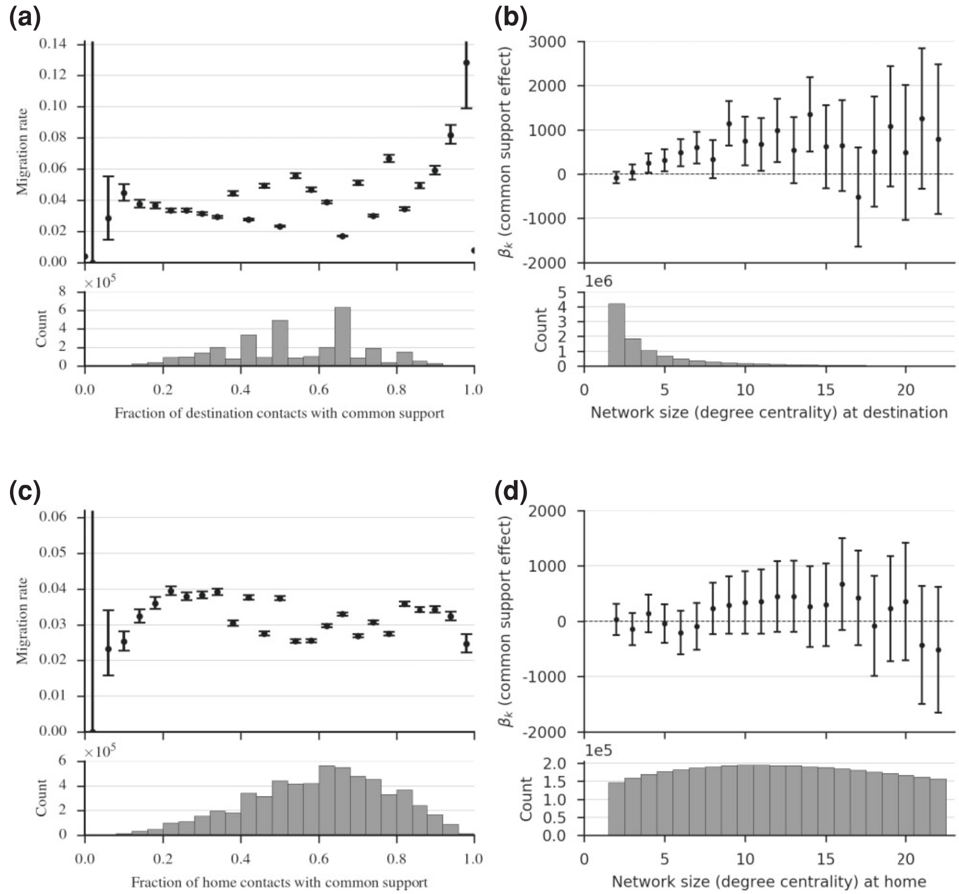

FIGURE 6

Migration and network "interconnectedness" (friends with common support). (a) Network support at destination. (b) Network support at destination, by degree. (c) Network support at home. (d) Network support at home, by degree

Notes: Network support indicates the fraction of contacts supported by a common contact (see Section 2.1). In all figures, the lower histogram shows the unconditional distribution of the listed variable. Figures in the left column (a and c) show the average migration rate for different levels of network support. Figures in the right column show the β_k values estimated with model (5), that is, the correlation between migration and support for individuals with different sized networks (network degree) after conditioning on fixed effects. Top row (a and b) characterizes the destination network; bottom row (c and d) characterizes the home network. Error bars for (a) and (c) indicate 95% confidence intervals, using the Wilson Score interval for binomial proportions. Error bars for (b) and (d) indicate 95% confidence intervals, two-way clustered by individual and by home–destination–month. Coefficients and standard errors on (b) and (d) are multiplied by 1,000 to make figures legible.

has more direct contacts in the destination, the value of interconnections between those contacts increases. Supplementary Figure A6 finds qualitatively similar results when using network clustering, instead of network support, as a measure of interconnectedness.[23]

23. The distinction between support and clustering is that the former counts the proportion of i's friends with one or more friends in common, the latter counts the proportion of all possible common friendships that exist—see Jackson (2010).

To provide further intuition for these results, and the variation that identifies them, we conduct the following test: We pull a random sample of 20,000 individuals who have exactly two contacts in a specific district for 4 consecutive months. We then calculate, for each person, whether those two contacts are became more connected or disconnected at the end of the 4-month period (by regressing a dummy for triadic closure on a linear time trend); we then compare the migration rate in month 5 among the population whose two contacts became more connected relative to the migration rate in month 5 of the population whose two contacts became less connected. The migration rate is 2.2% in the former group and 1.3% in the latter. In other words, when focusing on a sample who consistently have exactly two contacts in the destination, rates of migration are higher when a given individual's two contacts become more connected (over the 4-month period) than when they become more disconnected (over the 4-month period).

4.2.2. Network "extensiveness". The relationship between migration and network extensiveness is more surprising and subtle. Here, we are interested in the generalized comparison between G_1 and G_3 in Figure 1, and use the size of an individual's second-degree network $|D_i^2(G)|$ (*i.e.* their unique friends of friends) as a measure of extensiveness. Without controlling for the size of an individual's network, there is a strong positive relationship between migration and extensiveness in the destination (Figure 7(a)), and a strong negative relationship with extensiveness in the origin (Figure 7(c)). The shape of these curves resemble the relationship between migration rate and degree centrality shown earlier in Figure 5: average migration rates increase roughly linearly with the number of friends of friends in the destination, and decrease monotonically with the number of friends of friends at home.

Of course, the number of friends of friends a person has is heavily influenced by the number of friends that person has. Thus, Figure 7(b) and (d) shows how the number of friends of friends relates to migration, using specification (5) to hold fixed the number of friends. For the home network, Figure 7(d) indicates the expected pattern: the fact that all of the coefficients are positive suggests that given a fixed number of friends at home, people are less likely to leave when those friends have more friends. We also observe that the number of friends of friends at home matters more for people with fewer direct contacts—by the time an individual has a very large number of direct home contacts, their contacts' contacts matter less.

The surprising result is Figure 7(b), which indicates that the likelihood of migrating does not generally increase with the number of friends of friends in the destination, after conditioning on the number of friends. The friend of friend effect is positive for people with just one contact, but negative for people with three or more destination contacts. Averaged over all migrants, this effect is small but negative and statistically significant (row 3 of Table 2). This result is difficult to reconcile with standard models of information diffusion (*e.g.* Kempe *et al.*, 2003; Banerjee *et al.*, 2013). Indeed, much of the literature on migration and social networks suggests that, all else equal, individuals would be more likely to migrate if they have friends with many friends, as such networks would provide more natural conduits for information about job opportunities and the like.

We run a large number of empirical tests to convince ourselves that this pattern is not an artefact of our estimation or measurement strategy—several of these are described in Section 4.4. However, the data consistently indicate that the average migrant is no more likely to go to places where she has a large number of friends of friends. This is perhaps most transparent in Supplementary Figure A7, which shows the distribution of the count of friends of friends for all migrants and non-migrants with exactly ten friends in the potential destination. Among this sample, it is apparent that, on average, non-migrants have more friends of friends in the destination networks than migrants.

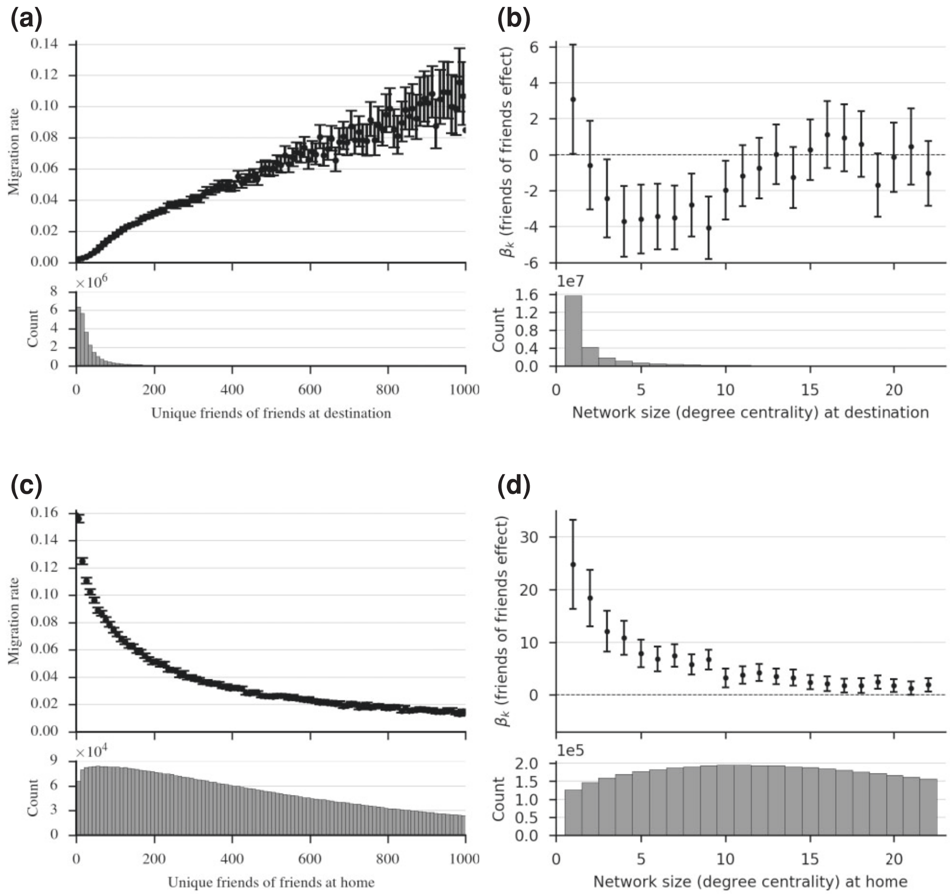

FIGURE 7

Relationship between migration and “extensiveness” (unique friends of friends). (a) Friends of friends at destination. (b) Friends of friends at destination, by degree. (c) Friends of friends at home. (d) Friends of friends at home, by degree

Notes: Main figures in the left column (a and c) show the average migration rate for people with different numbers of unique friends of friends. Figures in the right column show the β_k values estimated with model (5), that is, the correlation between migration and unique friends of friends for individuals with different numbers of friends, after conditioning on fixed effects. Top row (a and b) characterizes the destination network; bottom row (c and d) characterizes the home network. Lower histograms show the unconditional distribution of the independent variable. Error bars for (a) and (c) indicate 95% confidence intervals, using the Wilson Score interval for binomial proportions. Error bars for (b) and (d) indicate 95% confidence intervals, two-way clustered by individual and by home–destination–month. Coefficients and standard errors on (b) and (d) are multiplied by 1,000 to make figures legible.

4.3. *Heterogeneity and the “friend of friend” effect*

The effect that networks have on the “average migrant” masks considerable heterogeneity in how different types of migrants are influenced by their social networks. Supplementary Tables A3–A5 disaggregate the results from Table 2 along several dimensions that are salient in the migration literature: whether the migrant has previously migrated to the destination (Supplementary Table A3); whether the migrant stays in the destination for a long period of time (Supplementary Table A4); and whether the migration is between adjacent districts or over longer distances (Supplementary Table A5).

4.3.1. Heterogeneity and unawareness of the broader network. Several patterns can be discerned from these tables, but we focus our attention on how the network "extensiveness" effect changes with these different subgroups, as that was the most unexpected of the above results. Here, we find that for certain types of migration—repeat migrations, long-term migrations, and short-distance migrations—the number of friends of friends is positively or insignificantly correlated with migration rates. Each of these types of migration are significantly less common than the typical migration event (a first-time, short-term, and long-distance migration)—hence the negative average effect observed in Table 2.

This heterogeneity suggests one possible explanation for the unexpected "friend of friend" result: the average migrant may simply be unaware of the higher-order structure of their destination network. Such an explanation is supported by several other studies that find that people have incomplete information about the friends of their friends (Friedkin, 1983; Casciaro, 1998; Chandrasekhar *et al.*, 2016). This lack of information is likely to be most pronounced when the would-be migrant lives far from, or has less experience with, the destination friend's community. And indeed, this is what the heterogeneity suggests: the migrants who are positively influenced by extensive destination networks are the migrants who seem likely to be more familiar with the structure of those networks. When the destination is more familiar, it begins to resemble the home network, where people have good information on (proxies for) their friends' centrality (Banerjee *et al.*, 2019).

4.3.2. Recent migrants, recent visits, and strong and weak ties. Supplementary Tables A6–A8 indicate that the "extensiveness" of a migrant's destination network does not increase their probability of migration, even after accounting for several other factors that likely matter in the decision to migrate. For instance, we observe that people are more likely to go to places where their contacts have recently migrated. Coefficient estimates in column (3) of Supplementary Table A6 indicate that knowing a contact who previously made a specific origin–destination migration increases the likelihood of the migrant choosing that destination by 2–2.5 times the amount as knowing anyone else in the destination. The effect is similar for recent migrants who arrived in the destination very recently (last month) as for recent migrants who arrived at any point prior.

Likewise, Supplementary Table A7 controls for a binary variable indicating whether i ever appeared in district d in the month prior to t. In column (2), a "prior visit" is defined as making or receiving a call or text message from a tower in d; in column (3), we only consider activity that occurs between 6pm and 7am, in an effort to capture overnight visits. There is a strong correlation between such visits and migration (the effect is roughly 12 times as large as the effect of an additional direct contact in the destination). Controlling for these in-person visits does not change the qualitative role that networks play in shaping migration, but it does noticeably attenuate the effect of destination network structure (*i.e.* the effect of direct contacts decreases by roughly 30% and the effect of support decreases by roughly 40%), suggesting the in-person experience might substitute for network connections. Controlling for in-person visits has little effect on the influence of home network structure.

Supplementary Table A8 disaggregates the effect of social network connections by the "strength" of the social tie, where we define a "strong" tie as a contact with whom the individual communicates five or more times in the reference month (this is equivalent to the 90th percentile of communication frequency). Here, and consistent with recent work by Giulietti *et al.* (2018), we find that both strong and weak ties matter in migration: the effect of a strong destination tie is 0–34% larger than that of a weak destination tie; at home, the effect of a strong tie is 150–200% larger than a weak tie.

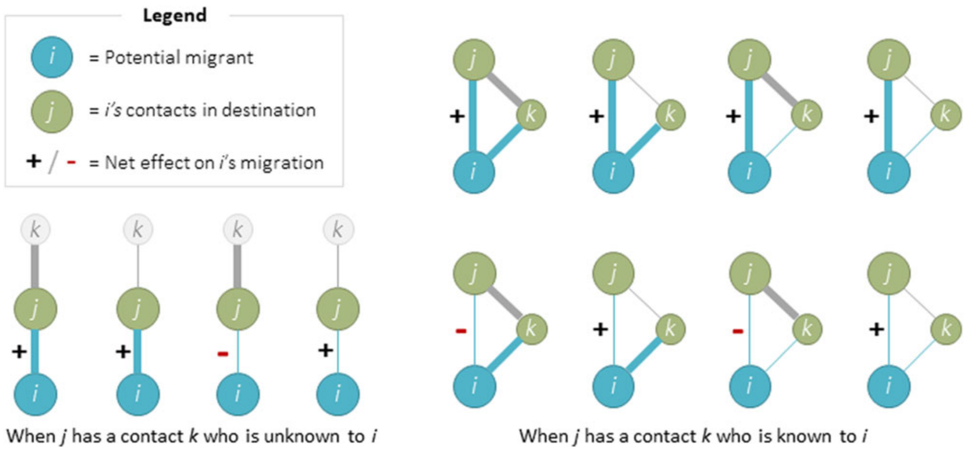

FIGURE 8
The role of (higher-order) strong and weak ties in a migrant's network

Notes: Thick edges represent "strong" ties and thin edges represent "weak ties" The ± signs summarize the effect that j has on i's likelihood of migration, based on the coefficients along the diagonal of Supplementary Tables A9 and A10.

Also interesting is the effect of *higher-order* tie strength on migration decisions. In particular, our main results suggest that a migrant i is drawn to locations where i's contact j has a friend in common k, but that i is indifferent or repelled if k is not a common friend of i. However, this average effect hides a more nuanced pattern: when disaggregating by tie strength, we observe that the negative effect is driven by situations where the i–j tie is weak but the j–k tie is strong—in other words, when the migrant has a tenuous connection to the destination and that tenuous connection has strong connections to other people in the destination.

These results are presented in Figure 8, which summarizes the regression coefficients from Supplementary Tables A9 and A10. The figure indicates the sign of the regression coefficient (using ±labels) from a regression of i's migration decision on the number of different types of i–j links, where type is determined by the strength of the i–j link (strong ties shown with thick lines, weak ties shown with thin lines) and the existence and strength of the j–k link. The four figures on the left, based on Supplementary Table A9, indicate that migrants are generally drawn to places where their contacts have many ties, but that they are deterred when their weak ties have a large number of strong ties. Similarly, the set of triangles on the right show all possible configurations of a supported i–j tie (based on Supplementary Table A10), and indicate that supported links are positively correlated with migration in all cases except when the i–j tie is weak and the j–k tie is strong.

This heterogeneity is consistent with the notion, proposed by Dunbar (1998) and others, that people might have a capacity constraint in the number of friendships they can effectively support, which in turn might induce a degree of rivalry for the attention of a friend. In our context, migrants may be drawn to places where they receive their friends' undivided attention.[24]

24. Dunbar originally proposed that humans could maintain roughly 150 stable relationships, since "the limit imposed by neocortical processing capacity is simply on the number of individuals with whom a stable inter-personal relationship can be maintained". In the migration context, Beaman (2012) and Dagnelie *et al.* (2019) find evidence that migrants may compete with each other for economic opportunities. See also Wahba and Zenou (2005), who empirically test the trade-off between information and rivalry in an Egyptian labour market survey. They show that up to a certain (network) size, the network information effect dominates the competition (rivalry) effect so that network is always beneficial for finding a job. However, above a certain size, the second effect dominates the first one so that agents have less chance of finding job when network size increases.

4.3.3. Beyond location-specific sub-networks. The regression results presented above calculate network extensiveness and interconnectedness based on location-specific sub-networks at home and in the destination. It is possible, however, that the social capital from network connections may cross geographic boundaries. For instance, a potential migrant i in home district h might receive information about a destination district d from a person k (who lives in d) via a common friend j that lives at home h or in a district other than d. We therefore show how results change when we relax restrictions on the location of the intermediate contact j.

Results in Supplementary Table A11 suggest that the main results—and in particular the negative role of extensiveness—do not depend on restrictions on the location of intermediate connections. For reference, column (1) of Supplementary Table A11 replicates the prior result from column (3) of Table 2. Column (2) of Supplementary Table A11 then shows results when i's direct contact j lives in the home district d; we observe that the coefficient associated with network extensiveness (*i.e.* unique friends of friends) remains negative, and is in fact much larger in magnitude than in column (1)—the increase is likely due to the fact that when i's home network includes an additional contact j to intermediate the connection to k, this also directly increases i's propensity to remain at home.[25] Column (3) allows for both types of network support (intermediated by destination friends and intermediated by home friends) to jointly influence the migration decision; both coefficients remain negative.

4.4. *Robustness and identification (revisited)*

Section 3.3 introduced our identification and estimation strategy. For our estimates to be causal requires the identifying assumption that $E[\epsilon_{idt}|\pi_{dt}, \mu_i, \eta_k] = 0$. In other words, we assume that the variation in higher-order network structure we observe is exogenous, conditional on the identity of the individual making the migration decision, the origin–destination–month choice being made, and the number of direct contacts the individual has in that destination in that month. Here, we discuss and test the limitations of that assumption, focusing on the two main threats to identification highlighted in Section 3.3: simultaneity and omitted variable bias.

4.4.1. Evidence against simultaneity: temporal lags and “shift-share” analysis. Our identification relies, in part, on the assumption that migrants do not strategically shape the higher-order structure of their social networks after making the decision to migrate. To support this assumption, Supplementary Figures A4 and A5 indicate that, even among eventual migrants, there are no sharp changes in average higher-order network structure in the months leading up to migration. Here, we provide additional analysis related to this identifying assumption.

First, we increase the lag between measurement of network structure and migration. Our main specifications (*e.g.* Table 2) test how migration decisions in month t (*e.g.* August 2008) relate to social network structure in month $t-2$ (*e.g.* June 2008). Supplementary Table A12 shows that results are qualitatively unchanged when migration in t is regressed on network structure in $t-6$ instead. Migrants may plan migrations more than 6 months in advance, but the similarity of the results using 2-month vs. 6-month lagged networks suggests that strategic network formation is not driving our results.

Second, we test a “shift-share” specification that relates migration decisions to changes in an individual's higher-order destination network structure, holding lower-order network structure

25. Likewise, the negative coefficient on destination support (intermediated by common friends at home) in column (2) is likely do to the fact that the additional contact at home required to increase destination support has a dampening effect on the migrant's propensity to leave home.

fixed. Specifically, we define an early period t_0 (*e.g.* 12 months prior to migration) and a late period t_1 (*e.g.* 2 months prior to migration), and measure the change in the higher-order network structure of each individual i between t_0 and t_1. In these specifications, we "freeze" the set of i's direct contacts at t_0, in an effort to reduce the endogenous decisions that i makes about their direct contacts (over whom they presumably have more direct control). The identifying variation in the shift-share specification comes from changes in the contacts of i's "frozen" contacts.

As intuition, Supplementary Figures A8 and A9 reconstruct Supplementary Figures A4 and A5, but hold fixed each migrant's contacts at $t-12$ to show how higher-order structure of that fixed network changes over time in the months leading up to, and immediately following, the eventual migration. In Supplementary Figure A8, we observe no sudden or gradual changes in the average number of friends of i's friends from $t-12$. In Supplementary Figure A9, we observe a gradual decrease in network support, but no sudden changes in the months immediately prior to migration.[26] The lack of sudden changes suggests that migrants are not systematically altering the high-order structure of their social networks in anticipation of migration.

Supplementary Table A13 presents regression results of the shift-share approach, in a format similar to Table 2, but now holding fixed i's contacts from t_0={6,12} months prior to migration. In both cases, we measure changes in higher-order network structure based on the connections of the intersection of i's contacts at t_0 and t_1, that is, the set of contacts who were connected to i in both t_0 and t_1. All specifications set the late period t_1 at 2 months prior to migration. The results of the shift-share analysis are broadly consistent with the main results in Table 2. Increases in friends of friends that occur within i's frozen-in-time contacts in the destination are insignificantly or negatively correlated with migration. Increases in support in the destination are positively associated with migration. We also note that the total predictive power of changes in network structure is limited (*i.e.* the partial R^2 values in Supplementary Table A13 are all less than 0.02). If migrants were systematically shaping their higher-order networks in anticipation of migration (and in advance of the $t-2$ lag used in our main specification), it is likely that such behaviour would better predict migration.

4.4.2. Evidence against omitted variable bias: increasingly restrictive fixed effects. Our preferred conditional logit specification (3) includes fixed effects for each individual (μ_i), each destination–month combination (π_{dt}), and each destination degree centrality (η_k). While these account for the most likely sources of omitted variable bias, there are scenarios in which this assumption could be violated (as in the carpenter/farmer example in Section 3.3). We therefore run a series of robustness checks that further isolate the identifying variation behind the regression results presented above.

In particular, we temporarily switch to a linear regression specification, which makes it possible to include a very restrictive set of fixed effect that we could not estimate with a conditional logit specification. In the linear specification, we define migration M_{ihdt} as a binary variable equal to 1 if the individual chooses to move from h to d at t and 0 otherwise, and estimate:

$$M_{ihdt} = \beta' Z_{ihd(t-s)} + \pi_{hdt} + \mu_i + \sum_k \eta_k \mathbb{1}(D_{id(t-s)} = k) + \epsilon_{ihdt}, \tag{6}$$

26. The gradual decrease in support is likely due to the fact that any specific edge in the network (including the j–k edge that provides support to the migrant i's edge with k) has a positive probability of disappearing over time. For each j–k edge that disappears, support will decrease unless it is replaced by a different j–m edge involving an m who is also connected to i. By contrast, we do not observe a gradual decrease in i's friends of friends. This is likely because j can replace a lost contact with any m—including those not connected to i—and maintain the same number of friends of friends.

where π_{hdt} are (home district $\times$ destination district $\times$ month) fixed effects, μ_i are individual fixed effects, and $D_{id(t-s)}$ is i's degree centrality.[27]

Column (1) of Supplementary Table A14 presents the results of estimating (6), using fixed effects similar to those in our preferred conditional logit specification (Table 2): π_{hdt}, μ_i, and η_k. The linear model results are qualitatively similar: while additional network support increases the likelihood of migration to the destination, additional friends of friends do not increase migration likelihood. Column (2) in Supplementary Table A14 then includes fixed effects for each *individual–month* pair, so that the identifying variation comes *within individuals in a given month* but across potential destination districts.[28] Column (3) instead includes separate fixed effects for each *individual–destination* pair, so that the β coefficients are identified solely by variation within individual–destination over time.[29] Column (4) includes fixed effects for each *individual–degree*, exploiting variation between all destinations where a single individual has the exact same number of contacts. Column (5), which includes over 600 million fixed effects, isolates variation within individual–home–destination observations over time. In all instances, the coefficients of interest are quite stable, and in particular, the average effect of additional friends of friends is either negative or insignificant (or both).

4.4.3. Additional tests of robustness. We perform several additional tests to check whether the main results are sensitive to different measurement strategies used to process the mobile phone data. Since these results show a very similar picture and are highly repetitive, we omit them for brevity:

- **How we define "migration" (choice of *k*)**: Our main specifications set $k=2$, that is, we say an individual has migrated if she spends 2 or more months in d and then 2 or more months in $d' \neq d$. We observe qualitatively similar results for $k=1$ and $k=3$.
- **How we define the "social network" (reciprocated edges)**: In constructing the social network from the mobile phone data, we normally consider an edge to exist between i and j if we observe one or more phone call between these individuals. As a robustness check, we take a more restrictive definition of social network and only include reciprocated edges, that is, cases where i calls j at least once and j calls i at least once.
- **Treatment of outliers (removing low- and high-degree individuals)**: Our main specifications remove from our sample all individuals (and calls made by individuals) with more than 200 unique contacts in a single month (this represents the 95th percentile). This is intended to remove spammers, calling centres, "public" phones, and large businesses. In robustness tests, we also remove individuals from our regressions with fewer than two contacts, to address concerns that individuals with just one or two friends could bias linear regression estimates, and that network support is sometimes considered undefined for individuals with fewer than two contacts.

27. When estimating equation (6), we include one observation for each potential destination d of each individual i in each month t. We define a "potential destination" as any non-home district $d \neq h$. Thus, the regression includes at most 26 observations for each individual i in each month t.

28. Such variation would occur if, for example, in a given month, a single migrant were choosing between two destination districts, had the same number of contacts in each district, and then decided to migrate to the district where his contacts were more interconnected—and if that additional interconnectedness exceeded the extent to which all networks in that destination were more interconnected.

29. This could reflect a scenario where an individual had been considering a move to a particular destination for several months, but only decided to migrate after his friends in the destination became friends with each other (the G_2 vs. G_1 comparison of Figure 1)—and where that tightening of his social network exceeds the average tightening of networks in that destination (as might occur around the holidays, for instance).

4.4.4. Summary. The fact that social networks are not randomly assigned makes it difficult to firmly establish the causal effect of networks on migration. In our setting, we exploit the rich data at our disposal to develop an identification and estimation strategy that offers, in our view, a plausible method to study the influence of higher-order network structure on migration. Specifically, we make the identifying assumption that $E[\epsilon_{idt}|\pi_{dt}, \mu_i, \eta_k] = 0$, and use the large quantity of data to estimate these fixed effects (π_{dt}, μ_i, η_k). This allows us to focus on how *higher-order* network structure, conditional on lower-order structure, relates to lagged migration decisions. The preceding sections provide evidence in support of this identifying assumption, and against many common alternative explanations for our results. Still, it remains an assumption, and we acknowledge that our identification is not bulletproof.

If this causal interpretation does not seem justified, the analysis nonetheless reveals a striking and hitherto undocumented relationship between social networks and migration. In particular, through all the robustness tests we have run, we consistently find that migrants are more likely to go to places where their social networks have certain types of higher-order structure. In particular, migrants are more likely to go to places where their contacts are interconnected. The presence of this positive correlation is accentuated by the fact that migrants are *not* more likely to migrate to places where their networks are more extensive, that is, where their friends have more unknown friends.

5. CONCLUSION

Social networks play a critical role in economic decision-making. This paper uses an extremely detailed dataset to understand how networks influence the decision to migrate. Our analysis suggests several new stylized facts about the relationship between social networks and migration. We find that migrants are consistently drawn to locations where their social networks are interconnected but that, perhaps most surprising, the average migrant is *not* drawn to places where their networks are extensive, that is, where their friends have lots of friends. Additional analysis suggests that this unexpected result may be due to the fact that migrants may have limited information about unknown destinations, and that they may feel competition for the attention of their well-connected friends. In additional, the granularity of our data allows us to document rich heterogeneity in how different types of migrants respond to social networks. For instance, we find that unlike the "average migrant," repeat migrants and long-term migrants are drawn to more extensive networks.

More broadly, these results highlight how new sources of digital data can provide nuanced insight into the role of social networks in consequential economic decisions. In contexts ranging from product adoption (Banerjee *et al.*, 2013) and disease transmission (Keeling and Eames, 2005) to the spread of new ideas and innovations (Rogers, 1962; Kitsak *et al.*, 2010), simple models of information diffusion have seen remarkable success. Such models often support the stylized narrative that the primary function of networks is to diffuse information about economic opportunities (cf. Rees, 1966; Ioannides and Loury, 2004). However, the patterns revealed by our data are hard to reconcile with these models, and suggest that some of the value of social networks comes from higher-order network interconnections. We thus hope that one broader contribution of this paper is to illustrate how, as rich social network data become available to researchers, those data can be used to test and distinguish between different models of network utility. Likewise, we expect that more nuanced empirical insights derived from such data can in turn help inspire advances in the theory of social networks.

Acknowledgments. We thank Oriana Bandiera, Lori Beaman, Francis Bloch, Thomas Chaney, Margherita Comola, Ben Faber, Marcel Fafchamps, Andrew Foster, Matthew Jackson, David Miller, Kaivan Munshi, Mark Rosenzweig, Sebastian Sotelo, Yves Zenou, as well as seminar participants at Stanford, Brown, Paris School of Economics, U.C. Berkeley, the BREAD/CEPR Conference on Development Economics, the NSF Conference on Network Science in Economics, the Economic Demography Workshop, and the Barcelona GSE Summer Forum for helpful feedback. We are grateful for financial support from GLM-LIC Grant C2-RA4-205 and UW RRF #65-7397. This research was approved by the U.C. Berkeley Committee for Protection of Human Subjects, protocol #042196-001.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

Data Availability Statement

Replication files for this paper are available at https://doi.org/10.5281/zenodo.10020030. This includes all of the analysis code required to replicate the tables and figures in this paper, as well as sample data files. The original data used in the paper were derived from mobile phone metadata obtained from a mobile phone operator in Rwanda. Due to privacy and confidentiality restrictions, these raw data cannot be shared publicly. Please see the replication files for instructions on how to access the original data.

REFERENCES

ALI, S. N. and MILLER, D. A. (2016), "Ostracism and Forgiveness", *American Economic Review*, **106**, 2329–2348.

AMBRUS, A., MOBIUS, M. and SZEIDL, A. (2015), "Consumption Risk-Sharing in Social Networks", *American Economic Review*, **104**, 149–182.

BALLESTER, C., CALVÓ-ARMENGOL, A. and ZENOU, Y. (2006), "Who's Who in Networks. Wanted: The Key Player", *Econometrica*, **74**, 1403–1417.

BANERJEE, A. V., BREZA, E. and CHANDRASEKHAR, A. (2012), "Come Play With Me: Eperimental Evidence of Information Diffusion About Rival Goods" (Working Paper, Stanford University).

BANERJEE, A., CHANDRASEKHAR, A. G., DUFLO, E., *et al.* (2013), "The Diffusion of Microfinance", *Science*, **341**, 1236498.

——, ——, —— (2019), "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials", *The Review of Economic Studies*, **86**, 2453–2490.

BANERJEE, A. V. and NEWMAN, A. F. (1998), "Information, the Dual Economy, and Development", *The Review of Economic Studies*, **65**, 631–653.

BARWICK, P. J., LIU, Y., PATACCHINI, E., *et al.* (2019), "Information, Mobile Communication, and Referral Effects" (Working Paper No. 25873, National Bureau of Economic Research).

BEAMAN, L. A. (2012), "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the US", *The Review of Economic Studies*, **79**, 128–161.

BEAMAN, L., BENYISHAY, A., MAGRUDER, J., *et al.* (2015), "Can Network Theory Based Targeting Increase Technology Adoption", *Unpublished Manuscript*.

BENGTSSON, L., LU, X., THORSON, A., *et al.* (2011), "Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti", *PLoS Med*, **8**, e1001083.

BERTOLI, S., FERNÁNDEZ-HUERTAS MORAGA, J. and ORTEGA, F. (2013), "Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions", *Journal of Development Economics*, **101**, 75–91.

BERTOLI, S. and RUYSSEN, I. (2018), "Networks and Migrants' Intended Destination", *Journal of Economic Geography*, **18**, 705–728.

BLUMENSTOCK, JOSHUA EAVN. (2012), "Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda", *Information Technology for Development*, **18**, 107–125.

BLUMENSTOCK, JOSHUA EVAN and EAGLE, N. (2010), "Mobile Divides: Gender, Socioeconomic Status, and Mobile Phone Use in Rwanda", in Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development (London: ACM) 6.

—— and —— (2012), "Divided We Call: Disparities in Access and Use of Mobile Phones in Rwanda", *Information Technology and International Development*, **8**, 1–16.

BORJAS, G. J. (1992), "Ethnic Capital and Intergenerational Mobility", *The Quarterly Journal of Economics*, **107**, 123–150.

BORJAS, G. J., BRONARS, S. G. and TREJO, S. J. (1992), "Self-Selection and Internal Migration in the United States", *Journal of Urban Economics*, **32**, 159–185.

BRAMOULLÉ, Y., KRANTON, R. and D'AMOURS, M. (2014), "Strategic Interaction and Networks", *American Economic Review*, **104**, 898–930.

BÜCHEL, K., EHRLICH, M. V., PUGA, D. and VILADECANS-MARSAL, E. (2020), "Calling from the Outside: The Role of Networks in Residential Mobility", *Journal of Urban Economics*, **119**, 103277.

BURT, R. S. (1992), *Structural Holes: The Social Structure of Competition* (Cambridge, MA: Harvard University Press).

CALVÓ-ARMENGOL, A. (2004), "Job Contact Networks", *Journal of Economic Theory*, **115**, 191–206.

CALVÓ-ARMENGOL, A. and JACKSON, M. O. (2004), "The Effects of Social Networks on Employment and Inequality", *The American Economic Review*, **94**, 426–454.

CARD, D. (2001), "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration", *Journal of Labor Economics*, **19**, 22–64.

CARLETTO, C., DE BRAUW, A. and BANERJEE, R. (2012), "Measuring Migration in Multi-topic Household Surveys" in Vargas-Silva, C. (ed) *Handbook of Research Methods in Migration*, chapter 10 (Edward Elgar Publishing).

CARRINGTON, W. J., DETRAGIACHE, E. and VISHWANATH, T. (1996), "Migration with Endogenous Moving Costs", *The American Economic Review*, **86**, 909–930.

CASCIARO, T. (1998), "Seeing Things Clearly: Social Structure, Personality, and Accuracy in Social Network Perception", *Social Networks*, **20**, 331–351.

CHANDRASEKHAR, A., BREZA, E. and TAHBAZ-SALEHI, A. (2016), "Seeing the Forest for the Trees? An Investigation of Network Knowledge" (Working Paper 24359, National Bureau of Economic Research February 2018).

CHANDRASEKHAR, A. G., KINNAN, C. and LARREGUY, H. (2018), "Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field", *American Economic Journal: Applied Economics*, **10**, 43–78.

CHI, G., LIN, F., CHI, G., *et al.* (2020), "A General Approach to Detecting Migration Events in Digital Trace Data", *PLoS One*, **15**, e0239408.

CHUANG, Y. and SCHECHTER, L. (2015), "Social Networks in Developing Countries", *Annual Review of Resource Economics*, **7**, 451–472.

COLEMAN, J. S. (1988), "Social Capital in the Creation of Human Capital", *American Journal of Sociology*, **94**, S95S120.

COMOLA, M. and MENDOLA, M. (2015), "The Formation of Migrant Networks", *Scandinavian Journal of Economics*, **117**, 592–618.

CORREIA, S., GUIMARÃES, P. and ZYLKIN, T. (2019), "Fast Poisson estimation with high-dimensional fixed effects", *The Stata Journal*, **20**, 95–115.

DAGNELIE, O., MAYDA, A. M. and MAYSTADT, J.-F. (2019), "The Labor Market Integration of Refugees in the United States: Do Entrepreneurs in the Network Help?", *European Economic Review*, **111**, 257–272.

DESHINGKAR, P. and GRIMM, S. (2005), *Internal Migration and Development: A Global Perspective* (New York: United Nations Publications).

DINKELMAN, T. and MARIOTTI, M. (2016), "The Long Run Effects of Labor Migration on Human Capital Formation in Communities of Origin" (Working Paper No. 22049, National Bureau of Economic Research).

DOLFIN, S. and GENICOT, G. (2010), "What Do Networks Do? The Role of Networks on Migration and "Coyote" Use", *Review of Development Economics*, **14**, 343–359.

DUNBAR, R. I. M. (1998), "The Social Brain Hypothesis", *Evolutionary Anthropology: Issues, News, and Reviews*, **6**, 178–190.

DUSTMANN, C., GLITZ, A., SCHÖNBERG, U., *et al.* (2016), "Referral-Based Job Search Networks", *The Review of Economic Studies*, **83**, 514–546.

EASLEY, D. and KLEINBERG, J. (2010), *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (**1st ed.**) (New York: Cambridge University Press).

EATON, J., KORTUM, S. S. and SOTELO, S. (2012), "International Trade: Linking Micro and Macro" (Working Paper No. 17864, National Bureau of Economic Research) Series: Working Paper Series.

EDIN, P.-A., FREDRIKSSON, P. and ÅSLUND, O. (2003), "Ethnic Enclaves and the Economic Success of Immigrants–Evidence from a Natural Experiment", *The Quarterly Journal of Economics*, **118**, 329–357.

FAFCHAMPS, M. and SHILPI, F. (2013), "Determinants of the Choice of Migration Destination", *Oxford Bulletin of Economics and Statistics*, **75**, 388–409.

FEWS NET Rwanda. (2014), "Rwanda Food Security Outlook: July to December 2014", Famine Early Warning Systems Network Food Security Update. https://fews.net/sites/default/files/documents/reports/Rwanda%20FSO%20July_December%202014.pdf.

FRIEDKIN, N. E. (1983), "Horizons of Observability and Limits of Informal Control in Organizations", *Social Forces*, **62**, 54–77.

GHOSH, P. and RAY, D. (1996), "Cooperation in Community Interaction Without Information Flows", *The Review of Economic Studies*, **63**, 491–519.

GIULIETTI, C., WAHBA, J. and ZENOU, Y. (2018), "Strong Versus Weak Ties in Migration", *European Economic Review*, **104**, 111–137.

GRANOVETTER, M. S. (1973), "The Strength of Weak Ties", *American Journal of Sociology*, **78**, 1360.

GREENWOOD, M. J. (1969), "An Analysis of the Determinants of Geographic Labor Mobility in the United States", *The Review of Economics and Statistics*, **51**, 189–194.

GUITERAS, R., LEVINSOHN, J. A. and MOBARAK, A. M. (2019), "Demand Estimation with Strategic Complementarities: Sanitation in Bangladesh" (CEPR Discussion Paper No. DP13498) https://ssrn.com/abstract=3328509.

HANSON, G. H. and WOODRUFF, C. (2003), "Emigration and Educational Attainment in Mexico" (Technical Report, University of California at San Diego).

HARE, D. (1999), "'Push' Versus 'Pull' Factors in Migration Outflows and Returns: Determinants of Migration Status and Spell Duration among China's Rural Population", *The Journal of Development Studies*, **35**, 45–72.

HU, Y. (2005), "Efficient, High-Quality Force-Directed Graph Drawing", *Mathematica Journal*, **10**, 37–71.

IOANNIDES, Y. M. and LOURY, L. D. (2004), "Job Information Networks, Neighborhood Effects, and Inequality", *Journal of Economic Literature*, **42**, 1056–1093.

JACKSON, M. O. (2010), *Social and Economic Networks* (Princeton, NJ: Princeton University Press).

—— (2020), "A Typology of Social Capital and Associated Network Measures", *Social Choice and Welfare*, **54**, 311–336.

JACKSON, M. O. and NEI, S. (2015), "Networks of Military Alliances, Wars, and International Trade", *Proceedings of the National Academy of Sciences*, **112**, 15277–15284.

JACKSON, M. O., RODRIGUEZ-BARRAQUER, T. and TAN, X. (2012), "Social Capital and Social Quilts: Network Patterns of Favor Exchange", *The American Economic Review*, **102**, 1857–1897.

JACKSON, M. O. and WOLINSKY, A. (1996), "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, **71**, 44–74.

JACKSON, M. O. and YARIV, L. (2010), "Diffusion, strategic interaction, and social structure", in Benhabib, J., Bisin, A. and Jackson, M. (eds) *Handbook of Social Economics* (North Holland Press).

JACKSON, M. O. and ZENOU, Y. (2015), "Games on Networks", *Handbook of Game Theory with Economic Applications*, **4**, 95–163.

KARLAN, D., MOBIUS, M., ROSENBLAT, T., *et al.* (2009), "Trust and Social Collateral", *The Quarterly Journal of Economics*, **124**, 1307–1361.

KEELING, M. J. and EAMES, K. T. D. (2005), "Networks and Epidemic Models", *Journal of The Royal Society Interface*, **2**, 295–307.

KEMPE, D., KLEINBERG, J. and TARDOS, È. (2003), "Maximizing the Spread of Influence Through a Social Network", in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '03* (New York, NY, USA: ACM) 137–146.

KERMACK, W. O. and MCKENDRICK, A. G. (1927), "A Contribution to the Mathematical Theory of Epidemics", *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **115**, 700–721.

KINNAN, C. (2019), "Distinguishing Barriers to Insurance in Thai Villages", *Journal of Human Resources*, **57**, 44–78.

KINNAN, C., WANG, S.-Y. and WANG, Y. (2018), "Access to Migration for Rural Households", *American Economic Journal: Applied Economics*, **10**, 79–119.

KITSAK, M., GALLOS, L. K., HAVLIN, S., *et al.* (2010), "Identification of Influential Spreaders in Complex Networks", *Nature Physics*, **6**, 888–893.

KÖNIG, M. D., ROHNER, D., THOENIG, M., *et al.* (2017), "Networks in Conflict: Theory and Evidence From the Great War of Africa", *Econometrica*, **85**, 1093–1132.

KRANTON, R. E. (1996), "The Formation of Cooperative Relationships", *The Journal of Law, Economics, and Organization*, **12**, 214–233.

LAI, S., ERBACH-SCHOENBERG, E. Z., PEZZULO, C., *et al.* (2019), "Exploring the use of Mobile Phone Data for National Migration Statistics, *Palgrave Communications*, **5**, 1–10.

LENORMAND, M., PICORNELL, M., CANTÚ-ROS, O. G., *et al.* (2014), "Cross-Checking Different Sources of Mobility Information", *PLoS One*, **9**, e105184.

LIGON, E. A. and SCHECHTER, L. (2012), "Motives for Sharing in Social Networks", *Journal of Development Economics*, **99**, 13–26.

LU, X., WRATHALL, D. J., SUNDSØY, P. R., *et al.* (2016), "Unveiling Hidden Migration and Mobility Patterns in Climate Stressed Regions: A Longitudinal Study of six Million Anonymous Mobile Phone Users in Bangladesh", *Global Environmental Change*, **38**, 1–7.

LUCAS, R. E. B. (1997), "Internal Migration in Developing Countries", *Handbook of Population and Family Economics*, **1**, 721.

—— (2015), "Chapter 26—African Migration", in Chiswick, B. R. and Miller, P. W. (eds) *Handbook of the Economics of International Migration, Vol. 1 of Handbook of the Economics of International Migration* (Netherlands: North-Holland) 1445–1596.

MAHAJAN, P. and YANG, D. (2017), "Taken by Storm: Hurricanes, Migrant Networks, and U.S. Immigration" (Working Paper No. 23756, National Bureau of Economic Research).

MCKENZIE, D. and RAPOPORT, H. (2010), "Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks", *Review of Economics and Statistics*, **92**, 811–821.

MCKENZIE, D. J. and SASIN, M. J. (2007), "Migration, Remittances, Poverty, and Human Capital: Conceptual and Empirical Challenges", SSRN Scholarly Paper ID 999482, Social Science Research Network, Rochester, NY.

MILLER, D. and TAN, X. (2018), "Seeking Relationship Support: Strategic Network Formation and Robust Cooperation" (Working Paper, University of Washington).

MONDERER, D. and SHAPLEY, L. S. (1996), "Potential Games", *Games and Economic Behavior*, **14**, 124–143.

MONTGOMERY, J. D. (1991), "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis", *The American Economic Review*, **81**, 1408–1418.

MORTEN, M. (2019), "Temporary Migration and Endogenous Risk Sharing in Village India", *Journal of Political Economy*, **127**, 1–46.

MUNSHI, K. (2003), "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market", *The Quarterly Journal of Economics*, **118**, 549–599.

—— (2014), "Community Networks and the Process of Development", *The Journal of Economic Perspectives*, **28**, 49–76.

MUNSHI, K. and ROSENZWEIG, M. (2016), “Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap”, *American Economic Review*, **106**, 46–98.

National Institute of Statistics of Rwanda (2012), “The Evolution of Poverty in Rwanda from 2000 to 2011: Results from the Household Surveys (EICV)” (Technical Report, Kigali, Rwanda).

— (2014), “Migration and Spatial Mobility” (Technical Report, Kigali, Rwanda).

PATEL, K. and VELLA, F. (2012), “Immigrant Networks and Their Implications for Occupational Choice and Wages”, *The Review of Economics and Statistics*, **95**, 1249–1277.

PINDOLIA, D. K., GARCIA, A. J., HUANG, Z., *et al.* (2014), “Quantifying Cross-Border Movements and Migrations for Guiding the Strategic Planning of Malaria Control and Elimination”, *Malaria Journal*, **13**, 169.

REES, A. (1966), “Information Networks in Labor Markets”, *The American Economic Review*, **56**, 559–566.

ROGERS, E. M. (1962), *Diffusion of Innovations* (New York: Simon and Schuster).

SIMMEL, G. (1950), *The Sociology of Georg Simmel* (Vol. 92892) (New York: Simon and Schuster).

SOTELO, S. (2019), “Practical Aspects of Implementing the Multinomial PML Estimator” (Technical Report, University of Michigan).

TODARO, M. P. (1980), “Internal Migration in Developing Countries: A Survey”, in Easterlin, R. (ed) *Population and Economic Change in Developing Countries* (Chicago: University of Chicago Press).

TOPA, G. (2001), “Social Interactions, Local Spillovers and Unemployment”, *The Review of Economic Studies*, **68**, 261–295.

UGANDER, J., BACKSTROM, L. and MARLOW, C. (2012), “Structural Diversity in Social Contagion”, *Proceedings of the National Academy of Sciences*, **109**, 201116502.

United Nations Population Division. (2013), “Cross-National Comparisons of Internal Migration: An Update on Global Patterns and Trends” (United Nations Technical Paper No. 2013/1).

WAHBA, J. and ZENOU, Y. (2005), “Density, Social Networks and job Search Methods: Theory and Application to Egypt”, *Journal of Development Economics*, **78**, 443–473.

WESOLOWSKI, A., BUCKEE, C. O., PINDOLIA, D. K., *et al.* (2013), “The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales”, *PLoS One*, **8**, e52971.

WESOLOWSKI, A., EAGLE, N., NOOR, A. M., *et al.* (2013), “The Impact of Biases in Mobile Phone Ownership on Estimates of Human Mobility”, *Journal of The Royal Society Interface*, **10**, 20120986.

WILLIAMS, N. E., THOMAS, T. and DUNBAR, M. (2013), “Measurement of Human Mobility Using Cell Phone Data: Developing Big Data for Demographic Science” (University of Washington CSSS Working Paper #137).

WINTERS, P., DE JANVRY, A. and SADOULET, E. (2001), “Family and Community Networks in Mexico-U.S. Migration”, *The Journal of Human Resources*, **36**, 159–184.

World Bank Group (2017), *Reshaping Urbanization in Rwanda: Economic and Spatial Trends and Proposals* (Washington, DC: World Bank).