Social Network Structure and the Radius of Risk Pooling

Daniel S. Putman*

February 6, 2024
Click Here For Most Current Version

Abstract

Informal risk sharing, an important coping strategy, is mediated by social networks. Does risk sharing extend beyond immediate friends and family? What is its radius? I examine the radius of risk sharing as a function of social network structure. To do so, I employ community detection—a tool imported from network science—and dyadic regression. I find evidence the radius of risk sharing extends beyond direct connections. Using data from a behavioral experiment in Colombia, I find that detected community co-membership and distance-2 connections (i.e., friends of friends) explain co-membership in experimental risk pooling groups. I address network sampling using a census of social network connections and transfers from a village in Tanzania, simulating the effect network sampling on results from similar specifications in the Tanzania data. These methods may benefit those who seek to understand the quality of risk sharing when risk sharing groups are loosely defined and illegible to outsiders.

Keywords: Risk Sharing, Network Formation, Community Detection, Sampled Networks

JEL Codes: D85, L14, O12, O17, Z13

*Corresponding Author. Postdoctoral Researcher, University of Pennsylvania Center for Social Norms and Behavioral Dynamics. Email: putmands@gmail.com). *Acknowledgements*: This paper has benefited immensely from conversations with Travis Lybbert, Ashish Shenoy, Laura Meinzen-Dick, Arman Rezaee, and Cathy Putman, as well as comments from several editors and anonymous reviewers. All errors herein are my own. *Disclosure statement*: No conflicts of interest to declare.

1 Introduction

Risk pervades the economic lives of the poor, determining the crops they plant, what jobs they take, the investments they make, and where they live (Banerjee and Duflo, 2007; Collins et al., 2010). This fact can lead to costly distortions in decision-making (Elbers et al., 2007; Karlan et al., 2014). Similarly, vulnerability to uncertainty itself reduces welfare in an *ex ante* sense (Ligon and Schechter, 2003). Despite this, formal financial markets that deal explicitly with risk, including insurance markets, are often missing for the poor (Mccord et al., 2007; Demirguc-Kunt et al., 2018). In the absence of formal insurance markets, informal risk pooling built on trust and reciprocity has long been studied as an important method of managing risk (Scott, 1976; Fafchamps and Lund, 2003; Karlan et al., 2009). These social motivations are powerful but limited tools to ensure cooperation in the face of information asymmetries. As the size and diversity of risk sharing groups grow, it often becomes more difficult to rely on trust, reciprocity, or monitoring (via social networks) to ensure that they function well (Fitzsimons et al., 2018; Jain, 2020).

Where early theory and empirical work placed the village as the radius of risk sharing (Townsend, 1994), recent empirical work pushed dyadic relationships to fore (Fafchamps and Lund, 2003). Still, there are many reasons why risk sharing might extend beyond the local network neighborhood, including intuitive mobilization or introduction (Small and Sukhu, 2016; Small, 2021), flows over networks (Bramoullé and Kranton, 2007; Bourlès et al., 2017), and pooling within formal or informal groups (Genicot and Ray, 2003; De Weerdt and Dercon, 2006; Ambrus et al., 2014). A larger radius matters considerably for the quality of insurance provided by informal groups and networks. Consider a simple income sharing arrangement where people share some of their income in a group: paying into the pot if their luck is good and receiving a payment if their luck is bad. The larger the group one can shares risk with, all else held equal, the lower the variability of the income shared—stabilizing the incomes of all involved in the arrangement. However, it is still unclear the radius and scale at which informal arrangements can work to share risk. A meaningful distance exists between direct friends and family and 'six degrees of separation' that might serve as the radius of a village (Strogatz and Watts, 1998; Henderson and Alam, 2022). Moreover, it is clear that this radius is a function of the network structure it is embedded within (Granovetter, 1985).

In this paper, I work to uncover the radius of risk sharing examining what embedded network structure best explains participation in risk sharing. To answer this question, I draw on *community detection* (Fortunato, 2010; Newman, 2012); dyadic regression, an econometric model of network formation; and network structures found to be relevant in past explorations of risk sharing. I argue that detected communities—clusters of individuals in networks who are closely connected within community and sparsely connected across communities—may serve as a principled method to determine the radius of risk pooling in social networks. In this way, the size of detected communities may also serve as an upper bound for the scale of risk sharing. To explore this hypothesis, I use data from a behavioral risk pooling experiment in Colombia.

Using data that combines social networks with risk sharing outcomes, I estimate econometric models of network formation to test whether social network structure can explain risk sharing behavior. I use dyadic regression, an approach which treats the dyad–any pair within the network (connected or unconnected)– as the unit of observation. I characterize network structure by quantifying how dyads are embedded within the network. More specifically, dyadic embeddings that might engender risk sharing include direct connections in social networks (Fafchamps and Lund, 2003), connections a step removed (i.e., 'friends of friends') (De Weerdt and Dercon, 2006; Bourlès et al., 2017), support (or the presence of a 'common friend,' sometimes called 'triadic closure') (Jackson et al., 2012), or more complex structures, like those quantified by detected communities. To detect communities I use the *Walktrap* algorithm (Pons and Latapy, 2005), which returns communities larger than the network neighborhood. This algorithm uses random walks over the network (i.e., from node-to-node, along edges) to understand where information or transactions might become 'trapped' within network structure. Intuitively, these same areas within the network might provide good information environments for risk sharing. In the first illustration I use data that combines real-world friends and family networks with experimental risk pooling (Attanasio et al., 2012b) to test whether social network structure can explain selection into these experimental risk pooling groups. This experiment is unique in that it allows for considerable coordination *ex ante* before risk sharing groups are formed, but no observable default *ex post*, meaning that networks serve only to limit adverse selection and not moral hazard.[1]

Estimates from the risk pooling experiment in Colombia indicate that detected community co-membership consistently helps explain co-membership in experimental risk pooling groups, i.e., even when controlling for other aspects of network structure known to be relevant. Shorter network distance translates into a higher propensity to share risk as well: Distance-1 and 2 connections consistently explain co-membership in experimental risk pooling groups, whereas distance-3 connections fail to do so. Furthermore, supported relationships, which capture tightly knit social network structure, are strongly correlated with co-membership in risk sharing groups. I argue that participants gather information about others in the network before deciding to join the same risk pooling group: they are successful in avoiding defaults, despite participant default being unobservable (Attanasio et al., 2012a). This process of gathering information may happen through introduction, (explicit or implicit) endorsement, and the ability to talk and interact with individuals during the experiment. For example, some indirect relationships may still reflect elements of bonding social capital, like a common group of contacts, which the community detection algorithm distills (Pons and Latapy, 2005). Since co-membership in groups with strategic defaulters would harm participants, participants maximize group size conditional on their ability to trust other participants. While most people who are unknown will be mistrusted (Sudarsky, 2018), those who can be befriended or endorsed through the network may be trusted or distrusted based on their words, actions, or discovery of common contacts, before groups are formed.

This paper's contribution to the literature is threefold. First, this paper contributes to the literature on risk sharing among the poor. In particular, the exercise helps to pin down—within this context—the radius of risk pooling. This adds value to a literature where models return radii of risk pooling which are anywhere from only those who are supported ties (Jackson et al., 2012) to the entirety of a village or administrative unit (Townsend, 1994). While Fafchamps and Lund (2003) and De Weerdt and Dercon (2006) (among others) show that risk sharing does not take place at the village level, this work refines their results. The fact that distance-2 and 3 connections and community co-membership matter for experimental risk pooling groups

[1]That is, the networks help in gathering information about what type of people one is joining a group with (the definition of *adverse selection*), but do not help in gathering information about the actions they take—particularly defaulting on loans (*moral hazard*).

implies a radius of risk pooling that extends beyond one's immediate network neighborhood. However, the (sometimes) failure of distance-3 nodes to explain risk sharing participation as well as the erosion of explanatory power over distance increases suggests that risk pools at a relatively more micro level than the village, consistent with their results. Thus, we can think of a true radius of risk pooling in this setting as occurring at a meso-level. Empirically documenting this meso-level radius of risk pooling adds credence to approaches that model risk pooling at the sub-village level in groups and/or networks, particularly theoretical work by Genicot and Ray (2003), Bloch et al. (2008), and Ambrus et al. (2014).

Second, it contributes to the broader social science literature on social network mobilization. In particular, I provide further evidence that people do not always 'map their network' before activating instrumental relationships (Small and Sukhu, 2016; Small, 2021) as they might do when answering survey questions about these networks. While other work focuses on the emotional and psychological dimensions to network mapping, I document this from the perspective of network structure. The extension of risk pooling beyond network neighborhood is at first counter-intuitive. Very often improvements in measurement simply make clear to the econometrician what the respondents or participants of the study already understand. For example, when we measure risk sharing networks (as opposed to villages), we elicit what respondents already know about their networks. However, measurement using community detection often pairs people who are not aware that they lie within each other's risk pooling groups. In this way, study participants may not fully appreciate the extent of their own risk pool beyond their network neighborhood until they activate these connections.

Third, this paper contributes to the literature on tie strength in risk-sharing networks by introducing community detection methods. It is well established that kinship, reciprocal, and supported ties have important relationships to tie strength (Granovetter, 1973; Fafchamps and Lund, 2003; Jackson et al., 2012). However, community co-membership yields a new and useful correlate of tie strength in networks, which I interpret similarly to the concept of support. Being co-members in a detected community suggests that one is in an environment where information flows freely, which aids in managing information asymmetries. Morevoer, community co-membership is not perfectly aligned with other measures of network distance explored here, which is to its a benefit in decomposing complex networks. This insight is most clear in the fully interacted results using the Colombia data, which yield multiple radii of risk sharing at different intensities. Supported relationships that also lie within communities serve as the strongest ties, with a 24 percentage point excess probability of joining the same risk sharing group, relative to an 8 percentage point excess probability among a similarly supported pair who are not co-members of a community.[2]

I address several threats to validity. Most critically, I test whether social network structure explains risk sharing transfers using detailed data from a village in Tanzania (De Weerdt, 2018). Since the Tanzania Nyakatoke Network forms a census of a village, I use a network sampling simulation using this data to address potential biases due to network sampling in the Colombia illustration and explore how community detection differs when networks are sampled. While sampling induces measurement error, dyadic relationships computed from sampled data are both strongly correlated with their census counterparts and—aside from detected communities—tend to produce similar regression estimates. For their part, this

[2] This is expressed as *excess* probability of matching as it is in excess of session level fixed effects.

exercise gives nuance to our interpretation of detected communities. In particular, while coefficients on community co-membership are not as stable, the results give credence to the idea that detected communities proxy for closer relationships as fewer nodes are sampled. To my knowledge, this is the first study to address measurement error from node sampling in a dyadic regression framework.[3]

The results are robust to a number of other empirical exercises. First, I repeat the Colombia analysis using the close friends and family network, which restricts friends or family to those dyads living in geographically proximate dwellings. Second, I include a battery of measures of affinity and differences between individuals that might drive co-participation in risk sharing. More specifically, I use the approach suggested by Fafchamps and Gubert (2007), controlling for the dyadic sums and differences of baseline characteristics. Third, I check the robustness of my preferred specifications. In particular, because I am estimating linear probability models (LPM), I check that my predictions lie within the unit interval—a sufficient condition to avoid bias and inconsistency with LPMs (Horrace and Oaxaca, 2006). For any specification where this is not the case, I re-estimate those specifications with logistic regression. None of these exercises meaningfully change the pattern or interpretation of results. Finally, I also replicate the analysis within the Tanzania data using a number of definitions of transfers. While these results depart from the Colombia analysis to some degree, they may do so for any number of reasons (see Appendix C.3.3).

These results are relevant for policy decisions, evaluation, and design. First, recent evidence has shown that financial technology may interact with social and financial networks. First, access to new financial technology may have the unintended consequence of eroding—or complementing—informal financial and economic relationships (Dupas et al., 2019; Dizon et al., 2019; Banerjee et al., 2022). Additionally, in cases where they complement financial technologies, risk sharing networks might impact their adoption—particularly index insurance (Mobarak and Rosenzweig, 2013; Lin et al., 2020; Berg et al., 2022). Thus, understanding the radius of risk sharing is important in measuring measuring the effects to overall financial health as financial technologies are adopted and to understanding adoption of insurance products themselves. Second, understanding the radius of risk pooling can also improve evaluation and design by improving our understanding of the radius of spillovers by non-market mechanisms enabling social scientists to identify treatment effects when network-mediated spillovers confound their estimation. Third, risk sharing transfers are often considered as a component of policy design. These include both spillovers of treatments and also to measure financial health.[4] However, if network activation is unknown to participants and respondents of these programs, this has implications for design. In the case of spillovers, knowledge of the radius of risk pooling would inform the radius of pass-on treatment and might be used for purposes of *ex ante* targeting of spillovers. In the case of assessments of financial health, this alerts us to a specific form of recall bias: respondents might have access to sources of funds that they do not consider and therefore do not report.

[3] There is some work on related but distinct problems. Most closely related is work from Chandrasekhar and Lewis (2016) and Smith and Moody (2013) that considers node sampling in relation to egocentric network statistics (as opposed to dyadic network statistics). There is also considerable work on cases where nodes are observed but some or all edges are unobserved. See for example, Breza et al. (2020).

[4] See for example Heifer International's Pay-it-Forward mechanism (Janzen et al., 2018) for a design which explicitly acknowledges spillovers. See Karlan and Brune (2017) for an instrument which uses questions about social networks to measure financial health.

2 The Radius of Risk Sharing in Groups and Networks

To situate the results on the radius of risk sharing within the literature of risk sharing, I briefly summarize literature on risk sharing in groups and networks. Early work on informal risk sharing focused on the village as the relevant group with whom risk is shared (Townsend, 1994). Complete risk sharing is a natural benchmark for the degree of risk sharing observed in villages.[5] In contrast, studies of risk sharing arrangements emphasize the social and economic relationships that serve to mediate risk sharing *ex post* (Fafchamps and Lund, 2003; De Weerdt and Dercon, 2006; Collins et al., 2010). Notably, evidence of information asymmetries and other imperfections in risk sharing arrangements abounds (Ligon, 1998; Kinnan, 2021).[6] However, it is still unclear the radius and scale at which informal arrangements can work to share risk.

The simplest approach to the radius of risk pooling is to assume the only people who matter in one's network are those with one shares direct connections. Of course, evidence abounds that these connections do matter (Fafchamps and Lund, 2003; De Weerdt and Dercon, 2006; Jack and Suri, 2014; Blumenstock et al., 2016). However, the risk pool could extend beyond the network neighborhood in several ways. In some theoretical models, all members of a network component share risk completely (Bramoullé and Kranton, 2007). In others, sharing is differentiate within the component by their network distances. The rationale for these higher distance connections may be because of network dynamics (e.g., friends of friends may be introduced), or flows on networks (e.g., a transfer from one person to the next influences other transfers) (Belhaj and Deroïan, 2012; Bourlès et al., 2017).[7] Other idiosyncratic reasons may also drive mobilization, including intuitive mobilization, alter driven mobilization, and homophily (McPherson et al., 2001; Fafchamps and Gubert, 2007; Small and Sukhu, 2016; Small, 2021).

In this vein, one could also use group memberships to explore risk sharing connections beyond the network neighborhood. Genicot and Ray (2003) explore the formation of such risk pooling groups with limited commitment. Groups which are stable (in the sense that they are self-enforcing) are bounded in size. The result of bounded size is mirrored empirically in Fitzsimons et al. (2018). Bloch et al. (2008) addresses these network structures in the context of limited commitment, examining the stability of risk sharing networks.[8] In this case, networks must act as conduits for risk sharing transfers and also for information. The authors find that certain network structures facilitate the spread information more than others, which in turn makes punishment of reneging more effective. Finally, Ambrus et al. (2014) build a theoretical model

[5]For example, Diamond (1967) models how contingent commodity markets can achieve optimal outcomes by completely smoothing idiosyncratic risk. More precisely, if a risk sharing arrangement approximates complete contingent commodity markets in a village, Pareto optimal allocations of consumption are achieved by competitive equilibrium. In the absence of information asymmetries or other market imperfections, informal risk sharing arrangements can be argued to resemble these contingent commodity markets.

[6]Many rationales have emerged to explain the failure of village economies to achieve complete risk sharing. These explanations include (but are not limited to) hidden income and assets (Cabrales et al., 2003), moral hazard (Delpierre et al., 2016; Jain, 2020), transaction costs (Jack and Suri, 2014), and limited commitment (Coate and Ravallion, 1993). All of these serve to place constraints on bilateral risk sharing, or risk pooling at the village level, through information asymmetries.

[7]From the perspective of data collection, it is also possible that friends of friends might help pick up unmeasured links when networks are dense and clustered.

[8]Notably these are exogenous networks for which stability is checked; this work does not explain the formation of the networks themselves.

of the effect of network structure on *ex post* consumption risk sharing. The authors find that commonly observed network structures do not imply complete risk sharing. Moreover, they hypothesize that in the case of incomplete risk sharing after the realization of shocks, risk sharing 'islands' will emerge where consumption is smoothed, resulting in good local risk sharing. These islands tend to feature a dense local network structure that is not well connected to other portions of the graph but is well connected within the island. Furthermore, risk sharing across islands is limited whereas risk sharing within islands is complete.

3 Community Structure in Risk Sharing Networks

How do we make larger structure legible in risk sharing network data? One approach would be to use labeled (and therefore formal or quasi-formal) groups.[9] While these are sometimes present, labeled, and legible to an econometrician, this is not always the case. Another approach might be to search for larger, complex features of networks. On approach might be to generalize from supported connections, one might consider cliques of nodes, in which all members of the clique are connected to all other members. These are likely related to risk sharing.[10] However, this leaves numerous questions unresolved: For example, should one treat cliques including a greater number of people included different than those with fewer? Likewise, what about an 'almost-clique,' missing just one relationship? Is it more natural to think of this as two cliques, or would we expect the two unconnected agents who have many friends in common provide insurance for each other?

We may be able to sidestep these issues entirely by using *community detection* algorithms to simplify the complex structure of networks (Newman, 2012). These communities consist of dense subnetworks within a larger network. Such community detection methods have been used in many contexts to identify the functional units within networks. Within the context of risk sharing networks, communities might help identify people who are likely to share risk beyond the network neighborhood. Furthermore, these communities yield a principled approach to simplifying complex networks in ways closely related to the theory of risk sharing in groups and networks. These densely connected groups should allow for ample opportunities for the flow of transfers and information. These properties are closely related to the properties of stable risk sharing networks from Bloch et al. (2008).[11] Communities also relate to the risk sharing islands seen in Ambrus et al. (2014). In particular, while risk sharing islands are *ex post* constructs, they share features with communities, including dense connections within the community or island, and few connections outside of the community or island. In this way, one might think of communities as *ex ante* areas of networks where one expects islands to form *ex post*.

Community structure is a latent feature of networks which must be uncovered using a community detection algorithm. My approach to uncovering these latent structures is based on random walks through

[9]For example, risk is sometimes shared explicitly in associations such as funeral insurance groups (Dercon et al., 2006).

[10]See Murgai et al. (2002) which uses an intuitive coding of clusters along irrigation canals in Pakistan and shows that insurance-related water exchanges in this context occur among households within these cliques.

[11]Network structures with high volumes of information pass-through include those with low density (such as trees and lines) and others with high density (such as the complete graph or a 'bridge' graph). 'Bridge' graphs, a set of two small cliques connected by one bridging link, are highly relevant here as they provide rationale for network structure that closely accords with community structure.

the network. In this setting, a random walker moves from node-to-node in the network by way of edges, randomly selecting the next node it visits among those in the network neighborhood. In particular, I use an algorithm proposed by Pons and Latapy (2005) and known as the *Walktrap* algorithm that uses random walks to estimate node similarity. The intuition for this method relies on the idea that within tightly knit sections of the network random walks become 'trapped' in the local network. Using a large number of random walks, the algorithm measures similarity between nodes and communities based on where these random walks land. If the walkers from two nodes (or two communities) tend to land on the same nodes, these two nodes can be thought of as close. I can then build communities using adjacent sets of nodes by restricting to those edges where the pairs of nodes are close in this sense.

The *Walktrap* algorithm proceeds as follows (Pons and Latapy, 2005):

1. Start with each node assigned into its own community. Compute distances for all adjacent communities in the network using random walks of length s (determined by the researcher).

2. Merge the two adjacent communities with the lowest distance between them into one community.

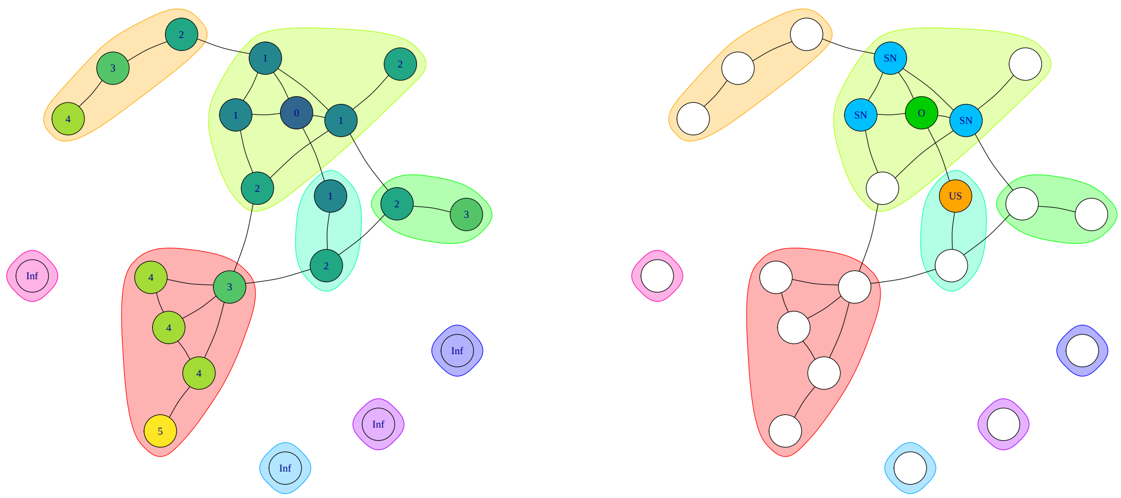
3. Recompute and update the distances between communities.

4. Repeat steps 2 and 3 until all communities have been merged into one community, recording each potential community assignment along the way.

5. This process yields a dendrogram (a hierarchical diagram documenting community merges). Using this dendrogram, I compare the modularity of all potential community assignments, and choose the one which has the highest modularity[12]

While several algorithms might mirror the intuition of risk sharing, I find Walktrap to have compelling features in this regard. In addition to mimicking the flow of goods or information on networks, they may generalize clustering in an interesting way. Suppose (against convention) one was limited to random walks of length one. This approach would consider nodes that featured common friends to be similar, correlating highly with clustering (and therefore support). Walks of length two would imply those dyads with the same friends of friends are similar. In this way, one can think of a supported relationship in a community to not only have the benefit of a common observer, but also to have a set of common observers at one step removed. All additional details of the community detection algorithm are included in Appendix A. Other algorithms, including the edge-betweenness based algorithm in Girvan and Newman (2004), result in similar community assignments.

[12]Modularity is the sum of connections above expectation occurring between individuals within a community. High modularity indicates that communities are dense, so I choose the community assignment with maximum modularity. See Appendix A.2 for more details about computation of this statistic.



(a) (b)

Figure 1: (a) Single session Colombia friends and family network with network distances and communities overlaid. Here 0 is the origin node, 1 indicates the set of distance-1 connections, 2 indicates the set of distance-2 connections, and so on. (b) Network support and detected communities overlaid in the same session network. Here 'O' is the origin, 'SN' indicates their set of supported neighbors, and 'US' their set of unsupported neighbors. Additionally, detected communities are represented by shaded regions in both visualizations.

4 Data and Context

4.1 The Risk Pooling Experiment

The data for the first illustration come from a laboratory experiment in Colombia and were obtained as replication files from Attanasio et al. (2012b).[13] In addition to experimental behavior, the data features real-world social networks and a rich set of demographic variables. In this section, I briefly explain the risk pooling experiment, sampling, and recruitment, as well as the real-world social networks survey measures. The experiment was conducted in 70 Colombian municipalities and elicited information about both risk preferences and risk pooling groups in two rounds of play. The first round of play consisted of a gamble choice game. This was followed by a luncheon where individuals were allowed to talk and form risk pooling groups to share their winnings from a second gamble choice game. Finally, individuals played a second gamble choice game and winnings were distributed according to the formed risk pooling groups.

The first round of the risk pooling experiment consisted of a version of the Binswanger (1980) gamble choice game. In this round the experimental participants chose one gamble from a list of six presented to

[13]Given concerns about replicability in modern economics, it is perhaps worthwhile to note that I am able to successfully replicate the results of Attanasio et al. (2012a) in a push-button replication. This paper studied assortative matching on risk preferences using this experimental data. While these results are closely related to those as they draw on the same data, to my knowledge, the only directly replicated exercise in this paper (aside from data descriptions) is specification (2) in Table 15.

them. As can be seen in Table 3, these gambles increase in both expected value and variance of payouts. While in the original study this was used as an indicator of risk aversion, here it serves purely to make income random. After choosing their gamble, participants played the gamble of their choice and received a voucher for their payout.

Round two of the experiment consisted of a second gamble choice game with the opportunity to pool risk. This time, before meeting with the experimenters, the participants were allowed to form risk pooling groups in which winnings would be pooled and shared equally, which would be declared before the second set of meetings took place.[14] During the meetings, participants were given the chance to privately withdraw from their groups after seeing the outcome of their gamble, a fact they were informed of before forming groups. In this case, when they withdraw, they forfeit their share of the group earnings but do not need to share any of their earnings with their former group. The remaining group members would pool their gambles and share these equally. Thus, each group member's earnings depend on the size and composition of the group after any withdrawal.

Of 122 municipalities surveyed to evaluate Colombia's national cash transfer program Familias en Acción, 70 municipalities were randomly drawn to participate in the experiment. About 60 households from each municipality were invited to an experimental session in their municipality. Households were selected from among families in the poorest sixth of the national population. Household members who attended were largely female as transfers were specifically targeted toward women.

Networks were collected on the day of the experiment by asking each participant in the experiment if they knew other participants and to clarify the nature of the relationship (family or friend). To the degree the session network is not the network of interest, this data collection strategy yields an important sampling issue for these networks, which I further explore in Section 6.3.1.

4.2 Variable Construction

In both empirical illustrations, nodes represent participants or survey respondents, and edges represent their direct relationships (e.g. friends or family).[15] For a given network g, let $N = \{1, \ldots, n\}$ be the set of nodes and $E = \{ij\}$ where $i, j \in N$ be the set of edges. The relationships between actors are represented in an $n \times n$ real-valued adjacency matrix $\mathbf{A} = \mathbf{A}(g)$ where $A_{ij} = \mathbf{1}(ij \in E)$.[16] In both empirical illustrations, the adjacency matrix is defined symmetrically: $A_{ij} = A_{ji}$ for all $i, j \in N$.

[14]Participants were given around an hour to an hour and a half (during lunch) to form their groups.

[15]For readers seeking a refresher on network notation: A graph $g = g(N, E)$ is a set of *nodes*, N, and an *edgelist* E containing *edges*. Nodes are sometimes referred to as vertices and edges are often referred to as arcs, links, ties, or connections. In this article I used relationships and connections to refer to a broader set of phenomena, as described below.

[16]Here, and throughout the text $\mathbf{1}(.)$ denotes the indicator function,

$$\mathbf{1}(\text{condition}) = \begin{cases} 1, & \text{if condition is true} \\ 0, & \text{if condition is false} \end{cases} . \tag{1}$$

4.2.1 Risk Pooling Groups

For the Colombia data, the outcome of interest is whether or not a dyad of individuals joined the same experimental risk pooling group. Being in a risk pooling group with the other member of the dyad is referred to as co-membership in the risk pooling group. Note that in this case groups are non-overlapping: each participant can only be in one group. Formally, for $i \in \text{Group}_i$ and $j \in \text{Group}_j$, I define $\text{Group}_{ij} = \mathbf{1}(\text{Group}_i = \text{Group}_j)$. I also have risk pooling groups in the Tanzania data; however, these are not outcomes since they are longstanding organizations meant to share risk related to funeral expenses and illness (Dercon et al., 2006). Co-membership in these groups is defined similarly to the Colombia illustration. However, in addition to being explanatory as opposed to outcomes, these groups differ in that membership can overlap among the groups. Therefore, I define an indicator for if there is any overlap in group membership at the dyad level. Formally, letting Groups_i be the set such that $\text{Group}_i \in \text{Groups}_i$, $\text{Any Group}_{ij} = \mathbf{1}(\exists \text{ Group} \in \text{Groups}_i \cap \text{Groups}_j)$

4.2.2 Social Networks

The explanatory variables of interest are constructed from the network survey data. In contrast to the experimental risk pooling groups or transfers, the constructed networks feature of *ex ante* friendships and family ties. Using the Colombia data, I define a network consisting of friends and family. Similar to the one in Attanasio et al. (2012a), this network is undirected. The network is unilaterally defined, meaning that if either respondent recognized friendship or kinship, then the network features an undirected link there even when the other did not reciprocally acknowledge that friendship or kinship.[17] Third, and in contrast to Attanasio et al. (2012a), which used only geographically proximate connections, the network is unrestricted by location. For the Tanzania data, I define the network to be undirected and unilateral as to create greater consistency with the Colombia illustration.[18]

4.2.3 Dyadic Relationships

For both illustrations, I start by forming an undirected and unweighted friends and family network graph, g. I say i and j are connected (or $ij \in g$), if either i recognizes j as a friend or family member or j recognizes i. For a graph g, I define the adjacency matrix of direct connections $A_{ij}(g) = \mathbf{1}(ij \in g)$. For distance-2 connections, I find all respondents that can be reached in two steps but are not direct connections. Formally, I define distance-2 connections as $A^2_{ij} = \mathbf{1}(\min \text{distance}(i,j) = 2)$ where distance is the number of steps when traveling over edges between the two nodes. Distance-3 connections are defined as any dyad with a shortest path of three, such that $A^3_{ij} = \mathbf{1}(\min \text{distance}(i,j) = 3)$. Figure 1(a) plots distances

[17] This is not a statement about the network formation process itself, since I lack the data to test this using the replication data (Attanasio et al., 2012b). For example, unilateral links (as they are reported within the data) might be so for a number of potential reasons. In particular, given the underlying directed network data, I could estimate whether the data generating process is best described by bilateral link formation, unilateral link formation, or 'desire to link' as defined within Comola and Fafchamps (2014).

[18] Again, this is not a statement about the network formation process itself. In fact, Comola and Fafchamps (2014) finds that this network is best understood as the result of household desire to link as opposed to unilateral or bilateral network formation. Given data to determine network formation in Colombia, an interesting replication of this work might use varying network definitions guided by these models of network formation.

from an origin node within a network for a single session. Finally, supported relationships are any dyad where the two members share a third friend in common. I define supported connections as $S_{ij} = \mathbf{1}(ij \in g \text{ and } \exists k \text{ such that } ij, jk \in g)$.[19] Figure 1(b) plots the supported connections of an origin node within a network for a single session.

4.2.4 Detected Communities

In addition to the above network variables, I propose an additional candidate measure based on community detection. Community detection splits households in the risk sharing network into discrete groups within villages based on network structure of the friends and family network. Each respondent is assigned to exactly one community, and all communities are composed of at least one respondent. Formally, for $i \in C_i$ and $j \in C_j$, I define community co-membership as $C_{ij} = \mathbf{1}(C_i = C_j)$. Figure 1 plots communities within a network for a single session in Colombia data while Figure 3 plots these networks in the Nyakatoke Network.

4.3 Summary Statistics

4.3.1 Group Formation

Participation varies by experimental session in the Colombia illustration (Attanasio et al., 2012a). On average, around 34 people attended each session, though this ranges from 9 to 87 participants in each session. 86.9% of participants chose to join a risk pooling group. These groups tended to be small, with an average of 4.6 members. Strategic default was relatively rare: 6.4% of participants defected from their group after winning their second-round gamble. From the perspective of dyadic relationships, about 10.6% of dyads are within an experimental risk pooling group (see Table 4).

4.3.2 Social Networks

A single session network from Colombia is visualized in Figure 1. Additionally, to describe overall network structure, details of the social network characteristics for the different social networks are presented in Table 5 in the Appendix. I find an average degree of 3.5 in the Colombia networks. Further description of the networks can be found in Appendix B.3.

4.3.3 Detected Communities

I detect communities using the *Walktrap* algorithm. A possible first step in using the algorithm is tuning the number of steps used in random walks. However, I use the default setting (walks of length four) for two reasons. First, this ensures consistency across empirical illustrations to ease comparison, and second, this removes researcher degrees of freedom which might lead to 'cherry picking' of results. For more on how walk length matters for detected communities, see Appendix A.3.

[19] All definitions here apply likewise to the close friends and family network, except I restrict the initial network graph to only friends and family who are geographically proximate.

Table 1: Characterizing Dyadic Relationships by Community Co-Membership

	Colombia Friends and Family			Tanzania Nyakatoke		
	Comm. Co-Membership			Comm. Co-Membership		
Prop.	Within	Between	All Dyads	Within	Between	All Dyads
All Distance-1	43.1	3.1	10.3	20.7	3.7	7.0
Supported	34.9	1.9	7.9	17.5	2.8	5.6
Unsupported	8.2	1.2	2.4	3.2	0.9	1.4
Distance-2	45.2	19.3	24.0	56.5	33.6	38.0
Distance-3	10.9	22.8	20.6	22.6	52.8	47.0
Distance-4+	0.8	54.8	45.1	0.2	9.9	8.1
Same Group(s)				15.4	10.2	11.2
Prop. Dyads	18.0	82.0	100.0	19.2	80.8	100.0

Notes: Dyadic relationship as a proportion of total dyads with the same community co-membership status. 'Within' indicate nodes lie within one community while 'between' indicates nodes within the dyad lie in separate communities. Supported dyads are directly connected with at least one additional node connected to both. Unsupported are connect with no other node connected both.

Results of community detection are presented alongside network characteristics in Table 5. In the Colombia networks, I find communities of average size 4.5 and modularity of 0.44.[20] The average communities yield approximately as many or more potential risk sharing partners than network neighborhood. In the case of the Colombia networks, communities provide about as many potential risk sharing partners as network neighborhood (around 3.5). While it is clear that distance-s connections extend to a radius beyond than the network neighborhood, this larger scale of communities *vis a vis* the network neighborhood suggests a larger radius as well. To further elucidate this point and to explore what kind of network structure these detected communities are capturing, in the next section, I summarize dyadic relationships within and between communities.

4.3.4 Dyadic Relationships

Summarizing dyadic relationships by community co-membership can help understand the network structure within communities and also the radius of risk sharing implied by these communities. Dyadic Relationships are summarized in Table 1. Dyads within communities have closer relationships than those across communities, both in terms of features like network distance and support. Nevertheless, detected communities would represent a wider radius of risk sharing than network neighborhoods, with a plurality (45.2%) of within dyads being distance-2 connections. Additionally, some dyads within communities were distance-3, though very few (less than 1%) are distance-4 or greater.

[20] Modularity is a measure of the quality of communities (see Appendix A.2).

5 Empirical Strategy

To test the explanatory power of various measures of the radius of risk pooling, I use dyadic regression, an econometric model of network formation. This method is easily interpretable and ideal for a first cut at a social network approach to answering this question. In these regressions, each pair of participants–whether connected or unconnected within the network–is treated as an observation. Based on the structure of the data, I only include dyads that were in the same session, since the networks were collected at the session level. Moreover, the approach here explicitly summarizes broader network structure at the dyad level to better take account of these complex dynamics.[21]

5.1 Main Specification

I start with a simple specification that seeks to explain participation in risk sharing using friendship or family ties in social networks. All dyadic regression specifications share a common outcome: participation in risk sharing, which I operationalize as co-membership in the experimental risk pooling group in the Colombia data and transfers in the Tanzania data.

If some additional measure is to add value above direct connections, it should be able to explain variation in participation in risk sharing. The first set of estimates focuses on three kinds of dyads: direct connections, supported connections, and co-membership in a detected community. A dyad is defined as a direct connection if both i and j recognize friendship or family ties. Second, this relationship is supported if there is a(t least one) additional respondent who is connected to both i and j. Finally, as the name suggests, a pair of respondents are co-members in a detected community if both belong to the same detected community (detailed descriptions of these variables are presented in section 4.2.) The main specification is as follows:

$$\text{Risk Sharing}_{ij} = \alpha_v + \beta_0 S_{ij} + \beta_1 A_{ij} + \gamma C_{ij} + \varepsilon_{ijv} \tag{2}$$

where α_v is a session fixed effect, Risk Sharing$_{ij}$ measures if i and j joined a group together, S_{ij} is an indicator equal to 1 if i and j have a supported connection, A_{ij} is an indicator equal to 1 if direct connection is present, and C_{ij} is an indicator equal to 1 if i and j are in the same detected community. Starting from the baseline that $\beta_1 > 0$, we want to test $\beta_0 > 0$ and $\gamma > 0$ conditional on the inclusion of A_{ij} in the regression. $\beta_0 > 0$ implies supported connections are more likely to participate in risk sharing together. On the other hand, $\gamma > 0$ indicates that detected communities explain risk sharing.

5.2 Longer Walks: Increasing the Radius of Risk Pooling

While detected communities may be one way we see increased radius of risk pooling, it may be that anyone within a specific radius is important for risk sharing. As depicted in Figure 1 and Table 1, there is an imperfect overlap between community co-membership and those who are proximate in networks. Here we care only about proximity, and not the community structure extracted from the networks. To test

[21] To increase the kinds of network motifs I account for would mean estimating Subgraph Generation Models (Chandrasekhar and Jackson, 2018). While such models would certainly add value, they are less easily interpretable models of network formation than dyadic regression would add considerable complexity to the analysis at hand.

this, I include dummies for those dyads who are 2 and 3 steps from each other. To test this, I include these indicators for 'longer walks' on their own as with measures of support and community. This specification can be written

$$\text{Risk Sharing}_{ij} = \alpha_v + \beta_0 S_{ij} + \sum_{s=1}^{3} \beta_s A_{ij}^s + \gamma C_{ij} + \varepsilon_{ijv} \quad (3)$$

where $A_{ij}^s = 1$ indicates that the shortest path between i and j is of length s. Here, I further test whether $\beta_s > 0$ for $s = 2, 3$. Similar to the previous tests of γ, tests of β_s might indicate that risk sharing extends beyond the network neighborhood. If rejected, these tests indicate that those further-flung members in an individual's social network are good candidates for pooling risk. However, since community co-membership and distance are closely related, the correlation when accounting for this measure is likely more meaningful. In terms of the magnitude of these effects, qualitatively, I would expect that closer dyads are more likely to match, i.e., $\beta_1 > \beta_2 > \beta_3 > 0$.

5.3 Estimation, Standard Errors, and Robustness

I estimate the above specifications using linear probability models (LPM) for purposes of interpretability and for the option to employ session level fixed effects. To ensure my results are robust to this specification choice, I examine the distribution of predicted probabilities. Specifically, if the LPM predictions lie on the unit interval, then estimates will not be biased by specification choice (Horrace and Oaxaca, 2006). Additionally, estimates from dyadic logistic regression are presented in Appendix F.2 as robustness checks.

I correct for non-independence of standard errors by clustering at the session level.[22] In both cases standard errors are robust to heteroskedasticity, which is of particular importance when estimating dyadic dependent LPMs.

I perform a number of other exercises to assess the validity of results, all of which I present in the Appendix. I re-estimate the LPM with a different social network, made up of geographically proximate friends and family. Third, while this paper is intended as a descriptive exercise, for both illustrations I include a large number of baseline characteristics as controls to understand the robustness of the relationship identified. All robustness checks can be found in Appendix F. Finally, for purposes of completeness, I replicate the analysis using the Tanzania data using transfers as outcomes. While these depart from the analysis in Colombia, this may be true for a number of reasons, as explained in the Appendix.

6 Results and Discussion

6.1 Network Structure and Experimental Risk Pooling in Colombia

6.1.1 Main Results: Support, Neighborhood, and Community

How well do these measures of the network explain co-membership in experimental risk pooling groups? Focusing on the first four specifications reported in Table 2, supported friends or family, friends or family,

[22] While I might also use dyadic-robust standard errors, cluster robust standard errors at a session level tends to be more conservativ (Fafchamps and Gubert, 2007; Cameron and Miller, 2014; Tabord-Meehan, 2019).

Table 2: Effects of Dyadic Relationships on Group Co-Membership: Colombia Friends and Family Network sans Controls

	Co-Membership in Risk Pooling Group						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.197***			0.0872***			0.0920***
	(10.15)			(4.19)			(4.52)
Distance-1		0.176***		0.0750***	0.193***	0.151***	0.0866***
		(11.02)		(4.92)	(12.00)	(9.43)	(4.81)
Distance-2					0.0388**	0.0187	0.0227
					(3.27)	(1.36)	(1.63)
Distance-3					0.00661	0.0000469	0.00203
					(0.64)	(0.00)	(0.18)
Same Community			0.115***	0.0583***		0.0527***	0.0488***
			(8.78)	(5.71)		(4.26)	(4.09)
N	88266	88266	88266	88266	88266	88266	88266
Session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses, standard errors clustered at the session level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and community co-membership enter positively and significantly (we always reject $\beta_0 = 0$, $\beta_1 = 0$, and $\gamma = 0$ at the 99.9% confidence level). However, the magnitudes of the estimates vary by specification. In particular, the three measures are strongly correlated, and may be picking up some overlapping information about network structure. Hence, I prefer to focus on specification (4), which includes all three variables. Here, being in the same community is associated with a 5.8 percentage point increase in the likelihood of joining the same risk pooling group, being in the network neighborhood is associated with a 7.5 percentage point increase in the likelihood of joining the same risk pooling group, and being in a supported relationship is associated with a 8.7 percentage point increase in the likelihood of joining the same risk pooling group.

First, this pattern of results confirms that those who are friends or family in social networks tend to pool risk together (Fafchamps and Lund, 2003; Fafchamps and Gubert, 2007). Unsurprisingly, the results related to distance-1 connections mirror those in Attanasio et al. (2012a), where they document that participants join the same experimental risk pooling group if they are close (geographically proximate) friends or family. Second, this reinforces that support drives risk pooling over and above network connections as might be suggested by Murgai et al. (2002) or Jackson et al. (2012). Finally, we see that detected communities drive risk sharing above and beyond these previously explored network measures.

6.1.2 Longer Walks: Distance-s Connections

While the basic measures of risk pooling seem to do well on their own and in concert, the same is not true distance-s connections. In specification (5) of Table 2, distance-2 connections enter significantly (at the 99.9% confidence level). However, the size of the association falls by roughly half with the inclusion of community dummies in specifications (6) and (7). Distance-3 connections, however, enter insignificantly across all specifications.

Taken together, the results from the main specifications and these start to tell a story of 'meso-level' risk pooling. If the most macro-level risk pooling as occurs at the village level (or the session level) and the most micro-level risk pooling occurs at the network neighborhood, the results here are at an intermediate level. Risk pooling falls far short of the diameter, or the longest minimum distance between two nodes of the largest components. However, it extends beyond the immediate network neighborhood. Those who are close are preferred as risk sharing partners—perhaps because they are better known—but there is some tolerance for network distance. Moreover, community co-membership does well in picking which individuals outside the network neighborhood are likely to be connected. This broader radius of risk sharing implies that risk sharing may be more able to smooth consumption than we would otherwise have imagined.

6.1.3 Group Formation and Adverse Selection

Why should network structure (and community structure in particular) matter in this experiment? Unlike the limited commitment frame which drives Ambrus et al. (2014) and Bloch et al. (2008), the experiment disallows *ex post* observation of default, making extrinsic motivation difficult to implement (i.e., the severing of links to punish the offender) (Attanasio et al., 2012a).[23] Despite this fact, the rate of strategic default is quite low (6.4%), suggesting that people tend to succeed in matching with trustworthy alters. Therefore, I interpret these results as a driven by the search for trustworthy risk sharing group members. While *ex ante* welfare (i.e., expected utility) should rise in group size when there is no threat of default it may fall when those who are not trustworthy enter the group.

Social trust is low in Colombia at the time of the experiment, but not extinct (Sudarsky, 2018). In particular, only 16% of men and 12% of women responded that most people can be trustedThe low trust environment motivates a key assumption: Those who are unknown via the network will not be trusted, unless more information can be gathered during the luncheon. That is, they are mistrusted until they are known, after which they may be trusted or distrusted. In this case, the network itself can serve to gather information about others in the network (e.g., through implicit endorsement) in addition to introductions and explicit endorsement. The relevant network for this information gathering activity is the one that is present on the day of the experiment because (1) it is salient and (2) the structure of the broader network is only imperfectly understood by participants (I return to these points in section 6.3.1 below).

In such a low trust environment, consent to enter risk sharing groups should serve as a key feature of the group formation process, meaning that group members should have veto power over others joining in

[23]Technically, the model in Ambrus et al. (2014) need not be limited commitment, but the set-up might be implied by such a model.

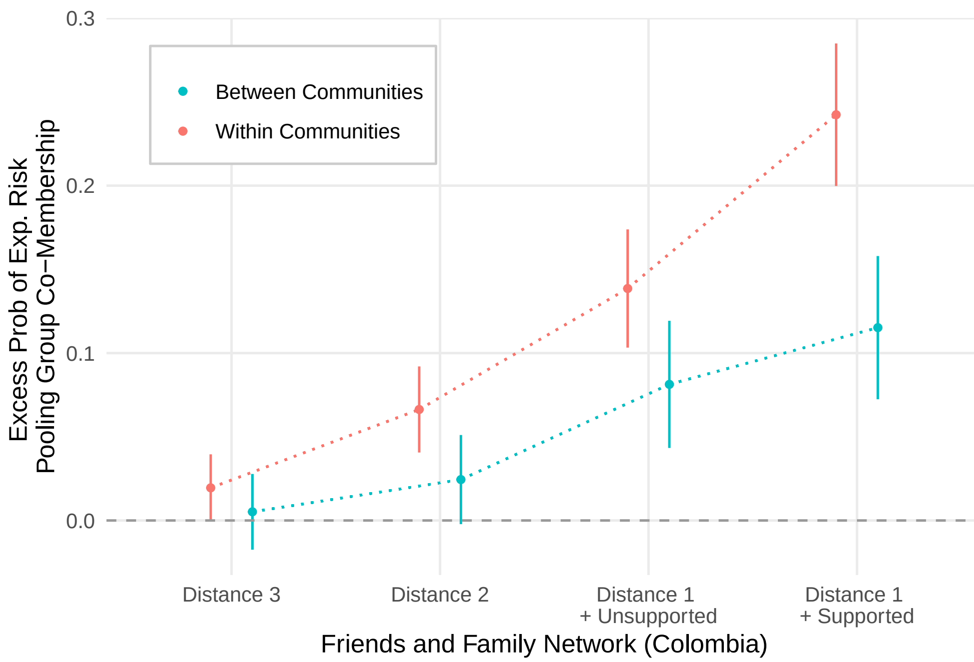


Figure 2: The Radii of Risk Sharing: Participation in risk sharing conditional on dyadic relationship featuring estimates from from a fully interacted model. Excess probability of co-membership in risk pooling group in Colombia friends and family network.

their group.[24] I appeal to an early model of coalition formation where simultaneous announcements are made (Hart and Kurz, 1983; Bloch and Dutta, 2011). In a simultaneous announcement game where participants cannot coordinate, one might worry about sparse coalitions. However, the experiment gives both considerable time for coordination and makes the need to coordinate is salient, allowing for participants to coordinate their responses in such a game (Attanasio et al., 2012a).

6.2 The Radii of Risk Sharing

A flexible, fully interacted regression specification allows us to inspect participation in risk sharing conditional on dyadic relationship. In doing so, picture of strong and weak ties in these networks emerges – multiple radii of risk pooling. Using estimates from the fully interacted specifications for co-membership in experimental risk pooling groups, any transfers, reciprocal transfers, and total transfers, I construct conditional expectations of participation in risk sharing and plot these in Figure 2.[25] In general, I find that the shorter the distance between two respondents, the higher the probability of co-membership regardless of community status. As seen in previous regression results, distance-3 ties tell us little about the probability of co-membership or the probability or size of transfers, whereas shorter distances are more informative.

[24]Consider a model with open group formation (free entry and exit), where those who are not trustworthy take part in the game. Those who are not trustworthy will default if their income is less than the expected average income of the group. However, if the number of each type are known, the gains to pooling risk fall, pushing out both trustworthy types and pushing the realization at which dishonest types will default, leading to an overall collapse in the group.

[25]The underlying estimates for these empirical specifications are presented in Table 13 and discussed in Appendix D.

As might be expected based on earlier results, co-membership in risk pooling group (using the Colombia friends and family network) shows a strong role for supported friendships as well as community co-membership. In Figure 2 not only does the probability of co-membership in a risk pooling group tend to increase the closer dyads are in terms of network distance, support, and community membership, but that community membership actually amplifies these other factors. Ties within detected communities are (weakly) better at explaining risk pooling group formation at every distance.[26]

6.3 Addressing Threats to Validity

6.3.1 Network Sampling and Measurement Error

The Colombia illustration may face a unique kind of measurement error, due to network sampling.[27] If the network observed among session members is not the salient network for the risk pooling experiment, then network sampling will induce some measurement error around the distance between nodes and the transitivity of connections (Smith and Moody, 2013). In the setting of dyadic regression, coefficients on distance-2, distance-3, and supported connections may be biased. In particular, distance-2 and 3 connections may not be recorded if intermediate connections are not sampled or may be recorded as being farther away from each other (e.g., distance-2 connections recorded as distance-3). Additionally, one may miss supporting nodes due to network sampling. This may lead to coefficients on distance-2, 3, and supported connections that are biased upward, or are upper bound estimates. It will also change the results of community detection, though it is more difficult to ascertain how community assignments might change in an unsampled network.

First, I argue that the effect of measurement error on our results may be lessened because the network collected in the experimental session may be the salient network for group formation within that narrow activity. That is, when assessing who to trust in forming experimental risk sharing groups network connections (even higher order ones like distance-2 connections) that are observable in the experimental session come to the fore in the process of endorsement and information gathering. Even if this social network is not the only network that is considered in forming risk pooling groups, the broader social network will be incompletely remembered, while the within session network is directly observable.[28] Furthermore, no information can be flow through the network via those who did not attend the session. Additionally, since there is not scope for *ex post* enforcement of risk sharing, those who did not attend do not impact the group formation in this way (for example, by spreading the fact they defaulted). Therefore, I argue that the salience of these within session network connections means the bias from network sampling should be limited.

Second, to address the issue of network sampling within the experimental data, I exploit the fact that the Tanzania Nyakatoke Network is a census of the village network. I use this data simulate a sampled

[26] Weakly stronger in the sense that I cannot always reject the null hypothesis that the means are equivalent.

[27] While this is conceptually related to Chandrasekhar and Lewis (2016), the implications differ due to the dyadic setting.

[28] There are important though limited examples where higher order network structure can be well recalled. See for example, Banerjee et al. (2019) in which members with high diffusion centrality are identified by asking who would be good at transmitting information. Notably, knowing highly central members is much different than knowing the relationships of less prominent members within a village.

network to understand the impact of this element on results from the dyadic regressions. The simulation proceeds as follows. Using the Nyakatoke Network, I randomly sample respondent households and keep only the network connections between those sampled households. I then proceed to generate a dyadic dataset of the remaining nodes and check the correlation between dyadic variables constructed from the sampled dataset and their census counterparts. Finally, I re-estimate regressions and compare average coefficients to their census estimates. This simulation yields three insights. First, in samples of 25% and 50% of the nodes I document (sometimes quite strong) correlations between dyadic relationships computed from sampled data and their census counterparts. Second, bias in regression coefficients is relatively small in practice and coefficients tend to be stable. Third, it gives nuance to our interpretation of communities. In particular, community detection tends to see greater bias than other variables, with estimated associations increasing in magnitude as sampling increases. While communities themselves are somewhat stable, the simulation indicates that detected communities proxy for closer relationships as fewer nodes are sampled. See Appendix F.1.1 for detailed methods and results. This would suggest that due to networks sampling in the Colombia data, communities also proxy for closer relationships, which might explain their greater explanatory power in this illustration. Additionally, while these results suggest small biases in estimates of other regressors from that illustration, estimates should be relatively robust to the issue of sampling.

6.3.2 Alternative Estimators, Networks, and Specifications

I test for robustness to network definition, using the close friends and family network from the Colombia data, as in Attanasio et al. (2012a). In this network the measure to *close* friends and family where closeness is defined as being geographically proximate. I find the results are robust to this alternative network, though coefficients are larger and somewhat nosier. See appendix F.1.2 for detailed results.

I test for robustness to specification choice, particularly the the choice to estimate coefficients using LPM. First, to diagnose if we should expect bias or inconsistency in coefficients due to this choice, I check the predicted outcomes for my preferred specifications. In specifications without fixed effects, predictions lie within the unit interval. Therefore, these specifications do not suffer from this specification choice (Horrace and Oaxaca, 2006). However, I find that when session fixed effects are employed, a small proportion of predictions lie outside the unit interval, suggesting that LPM may be biased or inconsistent. To assess this in practical terms, I present results from the LPM without session FE and dyadic logistic regression for each of my preferred specifications and compare marginal effects. I find that while marginal effects fall for my regressors of interest in both LPM and logistic specifications without FE, when I include session FE in my logistic specification, marginal effects are similar to those reported in LPM with FE. Therefore, I stand by my use of LPM with FE as they play an important role in controlling for common shocks at the municipality and session level. Full results of this analysis are presented in Appendix F.2.

6.3.3 Omitted Variable Bias

While this paper is descriptive in nature,it is instructive to assess omitted variable bias in order to understand the robustness of the relationship identified. For the Colombia illustration, there are two aspects of the setting to take advantage of which lessen these concerns. First, since the risk pooling experiment was conducted after real-world networks were realized, the results should not suffer from the possibility of reverse causality. Second, while common shocks may matter for both illustrations, I am able to control for these in estimates from the Colombia illustration using a fixed effects estimator. In particular, I include session level fixed effects in all regressions in the Colombia example to control for session-invariant features of group formation. These common shocks might include any variation in the execution of experimental protocols during the experiments or geographic heterogeneity.

To address concerns about omitted variable bias, I estimate all specifications from the main body of the paper (as well as some presented only in the appendix) using a selection-on-observables approach, controlling for dyadic characteristics that might drive co-participation in risk sharing. The estimates broadly accord with their counterparts. Appendix F.3 describes my variable selection approach and presents estimates.

6.3.4 Generalizability: Income, Rurality, and Gender

Another limitation of the results is that data used in the two empirical illustrations refer to specific segments of society. In particular, participants in Colombia tend to come from poor households in the rural areas targeted by the Familias in Accion program. In theory, networks in smaller scale settlements (like villages or municipalities) may yield more strong ties (for example, as a result of repeated interactions) compared to in larger settlements (of course as settlements become smaller yet, they may also bound network size).[29] Additionally, the sample in Colombia is predominantly female, which may limit the generalizability of the results. Furthermore, the strong skew towards one gender makes gender based quantitative comparison for gender based analysis difficult in the Colombia illustration.

7 Conclusion

Using dyadic regression, I test explanatory power of measures of network structure in explaining experimental risk-pooling outcomes. In doing this, I uncover how the radius of risk pooling depends on network structure. This allows me to correlate likely measures of risk sharing networks and groups with 'ground-truth' measures of risk sharing. Of the dyadic measures tested, three tend to be particularly useful in understanding the radius of risk pooling in the group formation experiment: direct connections, supported connections, and co-membership in detected communities. The third of these measures relies on community detection, a novel method to be applied to the study of risk pooling. In addition, distance-2 connections sometimes explain co-membership in experimental risk pooling groups, though these estimates are not as strong or stable as the other three.

These results point toward risk pooling that takes place at a level between the village (session) level and bilateral level. This understanding might guide how we think about the welfare derived from informal risk pooling. For example, one should be wary of any welfare calculations done under the assumption that *all* members of a village share risk. On the other hand, models that assume only bilateral risk sharing may

[29] This issue is not limited to the current work, considerable work on social networks draws on rural populations to ease the network boundary question (Laumann et al., 1983).

be conservative in this regard. When considering the literature on risk pooling, theoretical models that allow for this kind of 'meso-level' risk pooling become more intriguing, such as the work by Genicot and Ray (2003), Bloch et al. (2008), and Ambrus et al. (2014), among others.

New questions arise from community detection. If detected communities bound the radius of risk sharing, it becomes interesting how these communities are composed relative to network neighborhoods. In particular, it is often the case that network formation is guided by *homophily*, or the principal that 'birds of a feather flock together' (McPherson et al., 2001). Such homophily plays a strong role in risk-sharing networks in particular (Fafchamps and Gubert, 2007; Attanasio et al., 2012a; Barr et al., 2012). Are communities homophilous to the same degree as network neighborhoods? Interesting applications of communities include the study assortative matching on risk preferences (Putman, 2022). Finally, while risk sharing is an exciting application of community detection, it may prove valuable for places where social networks are relevant to the provision of goods. Questions still remain in the use of network science to understand risk sharing in networks as they might elsewhere. Where empirical tools match imperfectly with theory, clarifying the links between empirical modeling of these phenomena will be crucial.

References

A. Ambrus, M. Mobius, and A. Szeidl. Consumption Risk-sharing in Social Networks. *American Economic Review*, 104(1):149–182, 2014. doi: 10.1257/aer.104.1.149.

O. Attanasio, A. Barr, J. C. Cardenas, G. Genicot, and C. Meghir. Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, 4(2):134–167, 2012a. doi: 10.1257/app.4.2.134.

O. Attanasio, A. Barr, J. C. Cardenas, G. Genicot, and C. Meghir. Replication data for: Risk Pooling, Risk Preferences, and Social Networks, 2012b.

A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. *Review of Economic Studies*, 86(6):2453–2490, 2019. ISSN 1467937X. doi: 10.1093/restud/rdz008.

A. Banerjee, E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan. Changes in Social Network Structure in Response to Exposure to Formal Credit Markets, 2022.

A. V. Banerjee and E. Duflo. The Economic Lives of the Poor. *Journal of Economic Perspectives*, 21(1): 141–168, 2007. ISSN 0895-3309. doi: 10.2139/ssrn.942062.

A. Barr, M. Dekker, and M. Fafchamps. Who shares risk with whom under different enforcement mechanisms? *Economic Development and Cultural Change*, 60(4):677–706, 2012. doi: 10.1086/665599.

M. Belhaj and F. Deroïan. Risk taking under heterogenous revenue sharing. *Journal of Development Economics*, 98(2):192–202, 2012. doi: 10.1016/j.jdeveco.2011.07.003.

E. Berg, M. Blake, and K. Morsink. Risk sharing and the demand for insurance: Theory and experimental evidence from Ethiopia. *Journal of Economic Behavior & Organization*, 195:236–256, Mar. 2022. ISSN 01672681. doi: 10.1016/j.jebo.2021.12.035.

H. P. Binswanger. Attitudes toward Risk: Experimental Measurement In Rural India. *American Journal of Agricultural Economics*, 62(3):395–407, 1980.

F. Bloch and B. Dutta. Formation of networks and coalitions. In *Handbook of Social Economics*, volume 1, pages 729–779. 2011.

F. Bloch, G. Genicot, and D. Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143: 36–58, 2008. doi: 10.1016/j.jet.2008.01.008.

J. E. Blumenstock, N. Eagle, and M. Fafchamps. Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics*, 120:157–181, 2016. doi: 10.1016/j.jdeveco.2016.01.003.

R. Bourlès, Y. Bramoullé, and E. Perez-Richet. Altruism in Networks. *Econometrica*, 85(2):675–689, 2017. ISSN 0012-9682. doi: 10.3982/ecta13533.

Y. Bramoullé and R. Kranton. Risk-sharing networks. *Journal of Economic Behavior and Organization*, 64: 275–294, 2007. doi: 10.1016/j.jebo.2006.10.004.

E. Breza, A. G. Chandrasekhar, T. H. McCormick, and M. Pan. Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data. *American Economic Review*, 110(8):2454–2484, 2020. doi: 10.1257/aer.20170861.

A. Cabrales, A. Calvo-Armengol, and M. O. Jackson. La Crema: A case study of mutual fire insurance. *Journal of Political Economy*, 111(2):425–458, 2003. doi: 10.2139/ssrn.273401.

A. C. Cameron and D. L. Miller. Robust Inference for Dyadic Data, 2014.

A. G. Chandrasekhar and M. O. Jackson. A Network Formation Model Based on Subgraphs. Technical report, 2018.

A. G. Chandrasekhar and R. Lewis. Econometrics of Sampled Networks, 2016.

S. Coate and M. Ravallion. Reciprocity without commitment: Characterization and performance of information insurance arrangements. *Journal of Development Economics*, 40:1–24, 1993. doi: 10.1016/0304-3878(93)90102-S.

D. Collins, J. Morduch, S. Rutherford, and O. Ruthven. *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton University Press, 2010.

M. Comola and M. Fafchamps. Testing unilateral and bilateral link formation. *The Economic Journal*, 124 (579):954–976, 2014. doi: 10.1111/ecoj.12071.

M. Comola and M. Fafchamps. The missing transfers: Estimating misreporting in dyadic data. *Economic Development and Cultural Change*, 65(3):549–582, 2017. doi: 10.1086/690810.

J. De Weerdt. Risk-sharing and endogenous network formation, 2002.

J. De Weerdt. Nyakatoke Network Data, 2018.

J. De Weerdt and S. Dercon. Risk-sharing networks and insurance against illness. *Journal of Development Economics*, 81:337–356, 2006. doi: 10.1016/j.jdeveco.2005.06.009.

M. Delpierre, B. Verheyden, and S. Weynants. Is informal risk-sharing less effective for the poor? Risk externalities and moral hazard in mutual insurance. *Journal of Development Economics*, 118:282–297, 2016. doi: 10.1016/j.jdeveco.2015.09.003.

A. Demirguc-Kunt, L. Klapper, D. Singer, S. Ansar, and J. Hess. The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution. Technical report, 2018.

S. Dercon, J. De Weerdt, T. Bold, and A. Pankhurst. Group-based funeral insurance in Ethiopia and Tanzania. *World Development*, 34(4):685–703, 2006. doi: 10.1016/j.worlddev.2005.09.009.

P. A. Diamond. The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty. *American Economic Review*, 57(4):759–776, 1967. ISSN 00028282. doi: 10.1257/jep.6.3.79.

F. Dizon, E. Gong, and K. Jones. The Effect of Promoting Savings on Informal Risk-Sharing: Experimental Evidence from Vulnerable Women in Kenya. *Journal of Human Resources*, 55(3):963–998, 2019. doi: 10.3368/jhr.55.3.0917-9077R2.

P. Dupas, A. Keats, and J. Robinson. The effect of savings accounts on interpersonal financial relationships: Evidence from a field experiment in rural Kenya. *The Economic Journal*, 129(617):273–310, 2019. doi: 10.1111/ecoj.12553.

C. Elbers, J. W. Gunning, and B. Kinsey. Growth and risk: Methodology and micro evidence. *World Bank Economic Review*, 21(1):1–20, 2007. doi: 10.1093/wber/lhl008.

M. Fafchamps and F. Gubert. Risk sharing and network formation. *American Economic Review*, 97(2):75–79, 2007. ISSN 0002-8282. doi: 10.1257/aer.97.2.75.

M. Fafchamps and S. Lund. Risk-sharing networks in rural Philippines. *Journal of Development Economics*, 71:261–287, 2003. doi: 10.1016/S0304-3878(03)00029-4.

C. Fershtman and D. Persitz. Social Clubs and Social Networks. *American Economic Journal: Microeconomics*, 13(1):224–251, 2021. doi: 10.1257/mic.20180143.

E. Fitzsimons, B. Malde, and M. Vera-Hernández. Group Size and the Efficiency of Informal Risk Sharing. *The Economic Journal*, 128(612), 2018. doi: 10.1111/ecoj.12565.

S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010. ISSN 03701573. doi: 10.1016/j.physrep.2009.11.002.

G. Genicot and D. Ray. Group formation in risk-sharing arrangements. *Review of Economic Studies*, 70(1): 87–113, 2003. ISSN 00346527. doi: 10.1111/1467-937X.00238.

M. Girvan and M. E. J. Newman. Finding and evaluating community structure in networks. *Physical Review E*, 69, 2004. doi: 10.1103/PhysRevE.69.026113.

B. S. Graham. Dyadic regression. In *The Econometric Analysis of Network Data*, pages 23–40. 2020. doi: 10.1016/b978-0-12-811771-2.00008-0.

M. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1973. doi: 10.1017/S0269889712000130.

M. Granovetter. Economic Action and Social structure: The problem of Embeddedness. *American Journal of Sociology*, 91(3), 1985. ISSN 1676-5648. doi: 10.1590/S1676-56482007000100006.

S. Hart and M. Kurz. Endogenous Formation of Coalitions. *Econometrica*, 51(4):1047–1064, 1983.

H. Henderson and A. Alam. The structure of risk-sharing networks. *Empirical Economics*, 62(2):853–886, Feb. 2022. ISSN 0377-7332, 1435-8921. doi: 10.1007/s00181-021-02037-z.

W. C. Horrace and R. L. Oaxaca. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90(3):321–327, Mar. 2006. ISSN 01651765. doi: 10.1016/j.econlet.2005.08.024.

W. Jack and T. Suri. Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution. *American Economic Review*, 104(1):183–223, 2014. doi: 10.1257/aer.104.1.183.

M. O. Jackson, T. Rodriguez-Barraquer, and X. Tan. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1–45, 2012. doi: 10.1257/aer.102.5.1857.

P. Jain. Imperfect monitoring and informal insurance: The role of social ties. *Journal of Economic Behavior & Organization*, 180:241–256, Dec. 2020. ISSN 01672681. doi: 10.1016/j.jebo.2020.09.028.

S. Janzen, N. Magnan, S. Sharma, and W. Thompson. Short-Term Impacts of a Pay-It-Forward Livestock Transfer and Training Program in Nepal. *AEA Papers and Proceedings*, 108:422–25, 2018. doi: 10.1257/pandp.20181120.

D. Karlan and L. Brune. Measuring Global Financial Health. https://www.poverty-action.org/study/measuring-global-financial-health, 2017.

D. Karlan, M. Mobius, T. Rosenblat, and A. Szeidl. Trust and Social Collateral. *Quarterly Journal of Economics*, 124(3):1307–1361, 2009. doi: 10.1162/qjec.2009.124.3.1307.

D. Karlan, R. Osei, I. Osei-Akoto, and C. Udry. Agricultural Decisions After Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics*, 129(2):597–652, 2014. doi: 10.1093/qje/qju002.

C. Kinnan. Distinguishing barriers to insurance in Thai villages. *Journal of Human Resources*, 57(1):44–78, 2021. doi: 10.3368/jhr.57.1.0219-10067R.

E. O. Laumann, P. V. Marsden, and D. Prensky. The Boundary Specification Problem in Social Network Analysis. In R. S. Burt and M. Minor, editors, *Applied Network Analysis*. Sage Publications Ltd, 1983.

E. Ligon. Risk Sharing and Information in Village Economies. *Review of Economic Studies*, 65(4):847–864, 1998. doi: 10.1111/1467-937X.00071.

E. Ligon and L. Schechter. Measuring Vulnerability. *The Economic Journal*, 113(486):95–102, 2003.

W. Lin, J. Meng, and X. Weng. Formal insurance and informal risk sharing dynamics. *Journal of Economic Behavior & Organization*, 180:837–863, Dec. 2020. ISSN 01672681. doi: 10.1016/j.jebo.2019.04.016.

M. J. Mccord, J. Roth, and D. Liber. The Landscape of Microinsurance in the World's 100 Poorest Countries.

Technical Report April, The MicroInsurance Centre, 2007.

M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444, 2001. doi: 10.1146/annurev.soc.27.1.415.

A. M. Mobarak and M. R. Rosenzweig. Informal risk sharing, index insurance, and risk taking in developing countries. In *American Economic Review*, volume 103, 2013. ISBN 0002-8282. doi: 10.1257/aer.103.3.375.

R. Murgai, P. Winters, E. Sadoulet, and A. de Janvry. Localized and Incomplete Mutual Insurance. *Journal of Development Economics*, 67(2):245–274, 2002. doi: 10.1016/S0304-3878(01)00189-4.

M. E. J. Newman. Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1):25–31, 2012. doi: 10.1038/nphys2162.

P. Pons and M. Latapy. Computing communities in large networks using random walks. In *Lecture Notes in Computer Science*, volume 3733 of *International Symposium on Computer and Information Sciences*, pages 284–293. 2005. doi: 10.1007/11569596_31.

D. S. Putman. Sharing Covariate Risk in Networks: Theory and Evidence from Ghana, 2022.

J. C. Scott. *The Moral Economy of the Peasant*, volume 291. 1976. ISBN 0-300-02190-9. doi: 10.1007/s12115-014-9852-4.

M. L. Small. *On Mobilization: How Actors Mobilize Their Networks in Practice*, pages 573–595. Cambridge University Press, 1 edition, Sept. 2021. ISBN 978-1-108-87829-6 978-1-108-83997-6 978-1-108-81391-4. doi: 10.1017/9781108878296.042.

M. L. Small and C. Sukhu. Because they were there: Access, deliberation, and the mobilization of networks for support. *Social Networks*, 47:73–84, Oct. 2016. ISSN 03788733. doi: 10.1016/j.socnet.2016.05.002.

J. A. Smith and J. Moody. Structural effects of network sampling coverage I: Nodes missing at random. *Social Networks*, 35(4):652–668, 2013. doi: 10.1016/j.socnet.2013.09.003.

J. A. Smith, J. Moody, and J. H. Morgan. Network sampling coverage II: The effect of non-random missing data on network measurement. *Social Networks*, 48:78–99, 2017. doi: 10.1016/j.socnet.2016.04.005.

S. H. Strogatz and D. J. Watts. Collective Dynamics of 'Small-World' Networks. *Nature*, 393(June):440–442, 1998.

J. Sudarsky. WV5 Results Colombia. Technical report, 2018.

M. Tabord-Meehan. Inference With Dyadic Data: Asymptotic Behavior of the Dyadic-Robust t-Statistic. *Journal of Business & Economic Statistics*, 37(4):671–680, 2019. doi: 10.1080/07350015.2017.1409630.

J. C. Timoneda. Estimating group fixed effects in panel data with a binary dependent variable: How the LPM outperforms logistic regression in rare events data. *Social Science Research*, 93:102486, Jan. 2021. ISSN 0049089X. doi: 10.1016/j.ssresearch.2020.102486.

R. M. Townsend. Risk and Insurance In Village India. *Econometrica*, 62(May):539–591, 1994. doi: 10.2307/2951659.

A Community Detection

A.1 Distances

The Walktrap algorithm computes the similarity of nodes and communities using random walks. A random walker starts at a node i and moves to an adjacent node with probability equal to $1/d_i$ where d_i is the degree of i. This is repeated for s steps and the landing node k is recorded. For a given number of steps s (determined by the researcher), the distance between nodes i and j is defined (Pons and Latapy, 2005):

$$r_{ij}(s) = \sqrt{\sum_{k=1}^{n} \frac{(P_{ik}^s - P_{jk}^s)^2}{d_k}}. \tag{4}$$

P_{ik}^s is the probability that a walk starting at node i ends it's walk on node k. The distance overall can be thought of as the L^2 distance between P_{ik}^s and P_{jk}^s. Dividing by the degree of the receiving node helps control for the fact that these nodes will receive more random walks than others. Intuitively, nodes that send walkers to the same places in the network are close.

Building on this definition, they also define the distance between communities:

$$r_{C_1,C_2}(s) = \sqrt{\sum_{k=1}^{n} \frac{(P_{C_1,k}^s - P_{C_2,k}^s)^2}{d_k}}. \tag{5}$$

In this case, at the start of each random walk, the source within the community is drawn randomly and uniformly from members of that community:

$$P_{C,k}^s = \frac{1}{|C|} \sum_{i \in C} P_{ik}^s. \tag{6}$$

A.2 Modularity

The intuition behind modularity is that it measures dense connections within communities and sparse connections across communities. For this reason, it serves as a good validation tool to choose between possible community assignments. Newman (2012) defines modularity as follows: Let d_i and d_j be the degrees of nodes i and j, respectively. Let m be the number of edges in the graph. Suppose this graph was 'rewired,' breaking links into 'stubs' (i.e., half edges) and randomly reconnecting those stubs. The expected number of edges between nodes i and j from this rewiring is equal to $d_i d_j/(2m-1) \approx d_i d_j/2m$. If i and j are both high degree, it is more likely that they will be linked. Letting A_{ij} be the ijth entry of the adjacency matrix (defined $A_{ij} = \mathbf{1}(ij \in g)$), take the difference between these two numbers:

$$A_{ij} - \frac{d_i d_j}{2m}. \tag{7}$$

This difference is the observed connection (or lack of) less expected number of connections given the network degree of nodes i and j if community structure does not matter. Letting C_i be the community membership of node i, connections over expectation are weighted by the function C_{ij}, where $C_{ij} = \mathbf{1}(C_i = C_j)$. Finally, to express modularity, aggregate to the graph level and normalize by twice the number of links present.

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) C_{ij} \tag{8}$$

This serves as an easily computable and straightforward measure of the internal quality of communities. When all links occur within the communities, this statistic will be at its highest, reflecting a modular community structure.

A.3 Length of Random Walks for Walktrap

Performance (as measured by modularity) and community size differ by walk length. For the Colombia data, while modularity is relatively fixed. I find that walks of length two return communities with average size 3.93, whereas longer walks tend to return slightly larger communities on average. For example, walks of length five would return communities of 4.65. When used on the close friends and family networks, I see both smaller average community size and higher average modularity. For the Tanzania data, both community size and modularity depend more dramatically on the length of walks. Modularity increases with walk length as walks increase from length two to five. Likewise, community size increases from 4.96 to 13.22 members.

A.4 Detected Communities

In Tanzania, I find communities of size 10.8 and modularity of 0.30. Larger community size in Tanzania is likely related to higher average degree in the network data. Likewise, lower modularity is likely partially explained by lower density and clustering in this network. However, in the Nyakatoke network, communities provide an average additional 1.6 potential risk sharing partners (9.8 as compared to 8.2).

19.2% of dyads are within communities in Tanzania, in comparison to 18.0% of dyads in Colombia. In Tanzania, like in Colombia, dyads within communities have closer relationships than those across communities. The majority of within dyads in Tanzania (56.5%) are distance-2 connections. Comparing between Colombia and Tanzania, the closer connections within communities in Colombia is consistent with the greater density and clustering in these networks.

B Data Appendix

B.1 The Risk Pooling Game Incentive Structure

Letting $\ell \in \{1, \ldots, 6\}$ be an individual's type, earnings are equal to mean income from the gamble choice game. Neglecting withdrawal from the group, expected income from joining these risk sharing groups will be

$$E(y) = \sum_{\ell=1}^{6} q_\ell \times E(y_\ell) \tag{9}$$

where q_ℓ is the proportion of individuals who chose ℓ in the risk pooling group and $E(y_\ell)$ is the expected income of gamble ℓ. Likewise, the standard error of earnings will be $SD(\bar{y}) = \sqrt{Var(\bar{y})}$, where

$$Var(\bar{y}) = \frac{1}{N_G} \sum_{\ell=1}^{6} q_\ell^2 \times Var(y_\ell) \tag{10}$$

and N_G is group size. In the case where withdrawal is possible, it is rational for an individual to withdraw from the risk pooling group if their revealed income exceeds the expected income.

Table 3: Incentive Structure for the Gamble Choice Game

Gamble		Payoff Low	High	Expected Value	Standard Deviation
1	(safest)	3000	3000	3000	0
2		2700	5700	4200	2121
3		2400	7200	4800	3394
4		1800	9000	5400	5091
5		1000	11000	6000	7071
6	(riskiest)	0	12000	6000	8485

All amounts in Colombian pesos. Eeach gamble has a 50% probability of a low draw and a 50% probability of a high draw.

B.2 The Nyakatoke Network Data

The data for the second empirical illustration come from a detailed social network census in the community of Nyakatoke in Tanzania and were obtained as replication files from De Weerdt (2018). The data is chosen precisely because it is a census of households, which allows for investigation of issues of network sampling. The data includes rich panel data on household consumption and shocks as well as risk sharing transfers between all dyads in the network (De Weerdt and Dercon, 2006). In addition, individuals are asked to list those who they could personally rely on for assistance (De Weerdt, 2002), which forms the social network of interest.

Table 4: Summary of Dyadic Outcomes

Colombia	Obs	Mean/Prop	Std. Dev.	Min	Max
Risk Pooling Group Co-Membership (% Dyads)	88,266	10.6	30.8	0	1
Tanzania	**Obs**	**Mean/Prop**	**Std. Dev.**	**Min**	**Max**
Any Transfer (% Dyads)	7,021	14.6	35.3	0	1
Reciprocal Transfers (% Dyads)	7,021	5.6	23.1	0	1
Total Transfers (TZS)	7,021	326.8	1888.5	0	51650

B.3 Social Network Summary

The village network from Tanzania is presented in Figure 3. Notably, the Nyakatoke network is much larger than the average session network in Colombia, owing in part to the fact that it is a village census. This also relates to the average degree, or the number of (distance-1) connections each individual or household has. In particular, one would expect to see higher degree in a large network, since each node has more potential partners to name. Consistent with this intuition, I find an average degree of 8.2 in the Nyakatoke network compared to 3.5 in the Colombia network.

More details of overall network structure can be found in Table 5. Here I will summarize a number of network statistics, including density, clustering, and closeness.

Density is another measure of how how much social connection exists in a network and is computed by dividing the total number of connections by the maximum possible number of connections. In Colombia,

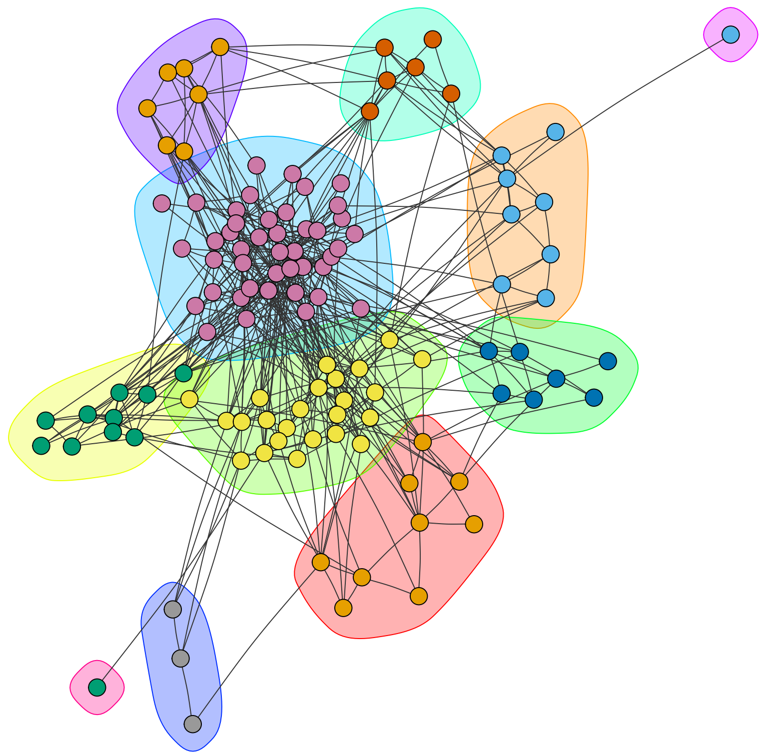


Figure 3: Nyakatoke Risk Sharing Network with Communities Overlaid. Detected communities are represented by both shaded regions and node color.

the average density of the friends and family network is 0.056, indicating that of all potential connections (within the session), only about 1 in 20 exist. Despite higher average degree, the Nyakatoke network is less dense, 0.035, indicating about 1 in 30 potential connections exist.

The clustering coefficient measures the transitivity of social connection. This is computed by dividing the number closed triplets of nodes by the number of all triplets, open or closed, where a triplet is a connected set of three nodes.[30] The Colombia networks feature an average clustering coefficient of 0.346. This indicates that network connections are transitive about one third of the time. The Nyakatoke network features a lower clustering coefficient of 0.188.

Finally, closeness measures the inverse of the average network distance between nodes. Higher values indicate closer networks while lower values indicate more distant networks. Closeness in the Colombia networks is 0.55, suggesting one can think of the average distance between nodes in a randomly chosen dyad (within one session) to be around 1.8 steps in these networks.[31] In the Nyakatoke network, closeness

[30] More formally, clustering coefficient answers the question: if ij and ik exists in the network what is the probability that jk is in the network as well?

[31] More precisely, closeness is computed by taking the average of the inverse of shortest path distance for nodes in the network. This particular definition is chosen to handle nodes in different components. When nodes are not in the same component, the shortest distance is often taken to be infinite, which would be problematic for any measure of distance. Therefore, I take closeness to be 0 as a convention. A value of closeness approaching one suggests that nodes are are rarely more than a step away from

is 0.436 indicating an average distance of 2.3 steps for a given dyad.

When compared to the friends and family network, the close friends and family networks are considerably less dense at the session level, but feature higher clustering coefficients. This is not unsurprising considering classic measures of bonding and bridging social capital. That is, we would expect more bonding (as opposed to bridging) relationships in the close friends and family networks relative to the unrestricted networks, where bonding is associated with support (Jackson et al., 2012).

Table 5: Characteristics of Social Networks and Detected Communities

	Colombia		Tanzania
Statistic	Friends and Family	Close Friends and Family	Nyakatoke Network
Nodes	33.971	33.971	119
	(11.954)	(11.954)	-
Edges	65.057	32.057	490
	(62.517)	(41.165)	-
(Average) Degree	3.520	1.677	8.235
	(2.477)	(1.483)	(4.991)
Density	0.056	0.026	0.035
	(0.044)	(0.022)	-
Clustering	0.336	0.425	0.188
	(0.202)	(0.301)	-
Closeness	0.547	0.742	0.436
	(0.169)	(0.162)	-
Number of Communities	11.457	18.100	11
	(7.152)	(7.791)	-
Community Size	4.515	2.134	10.818
	(3.511)	(1.018)	-
Modularity	0.44	0.59	0.30
	(0.17)	(0.20)	-

Standard errors in parentheses. For Colombia, both means and standard errors computed from session level statistics. For Tanzania, means and standard errors computed on the node level.

each other, on average. As closeness approaches zero, nodes are very far, or more likely in separate components.

C Network Structure and Risk Sharing Transfers in Tanzania

C.1 Summary

Turning to the Tanzania data, I replicate the analysis focusing on intra-village risk sharing transfers to measure participation in risk sharing.[32] Detected communities are larger in this illustration, reflecting more interconnected network data. I find that community co-membership helps to explain risk sharing data in addition to distance-1 connections across several transfer outcomes. However, I find that the association between community co-membership and transfers is crowded out by distance-2 and 3 connections.[33] While it is difficult to pin down every difference between the two illustrations I speculate that several mechanisms may play a role. First, the group formation process differs from the transfer process in important ways. For example, if group members holds veto power over entries, then group formation requires sufficiently strong ties and/or coordination between all members of the group (Hart and Kurz, 1983). Second, the information environment differs. While forms of moral hazard may join adverse selection as barriers to risk sharing, latitude for punishment also arises. Such 'self-enforcing' contracts might allow individuals to be less selective in who they choose to share risk with (Coate and Ravallion, 1993; Bloch et al., 2008).

C.2 Risk Sharing Transfers

The outcomes for the Tanzania data are constructed from risk sharing transfers. For each dyad, four measures of transfers are collected. Each member of the dyad is asked if they have given to the other member or received from the other member of the dyad, leading to reports which are sometimes discordant. This phenomenon is evaluated for the same dataset in Comola and Fafchamps (2017). Following their empirical results, which indicate there is likely under-reporting by one party when responses are discordant, I take the maximum reported flow from i to j and from j to i in each of the five rounds of data collection. Then I define three measures of transfers. The first outcome I define out of this data is an indicator variable for if any transfers are made within the dyad in any round. Second, I define an indicator variable for if these transfers were reciprocal. That is there was a transfer from i to j in some round and one from j to i in some (possibly other) round. Third, I define total transfers as the sum of all transfers in either direction across all rounds (reported in TZS). About 14.6% of dyads have made any transfers, though only 5.6% reciprocally. The average total transfers within dyad are 326.8 TZS, though this figure rises to 2238.4 TZS when considering only the 14.6% of dyads where transfers are made.

C.3 Main Results

C.3.1 Main Results: Support, Neighborhood, and Community

Using the Tanzania data, I test how network structure explains participation in risk sharing transfers in a real-world setting. Estimates of main results are presented in specifications (1)-(4) of Table 6. I correct for non-independence using dyadic-robust standard errors (Fafchamps and Gubert, 2007; Cameron and Miller, 2014; Tabord-Meehan, 2019). While support, distance-1 connections, and community co-membership all

[32] These transfers are risk sharing transfers as they were specifically elicited as coping strategies in the survey. See De Weerdt and Dercon (2006) for more.

[33] The detailed data from Tanzania also allow me to document the role of co-membership in pre-existing insurance groups. When further including co-membership in any of these insurance groups in specifications, two additional insights emerge. First, coefficients on co-membership in detected communities and co-membership in insurance groups have similar magnitudes, suggesting this informal network structure may be of similar importance to the formal structure of insurance groups. Second, the information from communities corresponds only weakly to that encompassed by the insurance groups, in that it explains transfers even when controlling for co-membership in insurance groups.

enter significantly when they are the sole explanatory variable, support is highly attenuated and insignificant when all three variables are included together. Likewise, community co-membership is highly attenuated, though still significant (at the 99% confidence level). As before, the three variables are highly correlated, which may explain some of the attenuation. Focusing on specification (4), I find being in the same community is associated with a 4.9 percentage point increase in engaging in any transfers, being in the network neighborhood is associated with a 63.0 percentage point increase in any transfers, and being in a supported relationship is associated with an imprecisely estimated 8.1 percentage point increase in the probability of joining the same risk pooling group.

Table 6: Effects of Dyadic Relationships on Transfers: Tanzania Nyakatoke Network sans Controls

	Any Transfers Within Dyad						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.722***			0.0805			0.0861
	(25.59)			(1.38)			(1.49)
Distance-1		0.714***		0.630***	0.787***	0.780***	0.712***
		(32.27)		(13.80)	(29.93)	(28.17)	(16.55)
Distance-2					0.141***	0.137***	0.138***
					(5.57)	(5.45)	(5.46)
Distance-3					0.0307***	0.0297***	0.0297***
					(7.10)	(6.94)	(6.95)
Same Community			0.167***	0.0487**		0.0125	0.0114
			(6.97)	(2.85)		(0.83)	(0.75)
N	14042	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.3.2 Longer Walks: Distance-s Connections

While with the exception of supported connections, the main results are similar to those from the Colombia data, results diverge more dramatically when I include distance-2 and distance-3 connections in the specification. In the final three specifications in Table 6, distance-2 and 3 connections enter positively and significantly with relatively similar magnitudes across specifications. Focusing on specification (7) which includes supported, distance-1, 2, and 3 connections, and community co-membership, I find being a distance-2 connection is associated with a 13.8 percentage point increase in making any transfer, and being a distance-3 connection is associated with a 3.0 percentage point increase in making any transfer. Likewise, the correlation between community co-membership and transfers is attenuated when included with distance-2 and 3 connections. It appears that social distance dominates who shares risk with whom in this village risk sharing network. One way to understand this fact is to recall the underlying network and community structure. In particular, others within communities are more distant in this illustration (see Table 1). In any case, it seemed that community detection is not able able to determine which distance-2 and 3 individuals were more relevant, as opposed to in the Colombia illustration.

C.3.3 Comparing Empirical Illustrations

Why do the results from Tanzania differ *vis a vis* the risk sharing experiment? In particular, while the formation of risk sharing groups was well explained by community co-membership, in the results from the Tanzania communities are crowded out by distance-2 and 3 connections. While I may not be able to pin down an exact answer to this, several mechanisms may play a role: the outcome in each illustration, the role of asymmetric information in that illustration, and how networks are collected.

1. *Groups vs. Transfers*: Risk sharing outcomes do not match one-to-one across these two illustrations. While in Colombia risk sharing groups are the outcome of interest, in Tanzania I observe dyadic transfers instead. Aspects of group formation might explain the difference in results. For example, a group formation processes with veto power (like that implicit in Hart and Kurz, 1983) requires sufficiently strong ties between all pairs of dyads in the group. It may be the case that being in the same community increases the likelihood of having more of these connections. In contrast, community structure may not matter as much for dyadic transfers because it collects information extraneous to that dyad.

2. *Asymmetric Information*: Working with the Tanzania Nyakatoke data, I do not have the same precision with which to understand risk sharing transfers as a function of asymmetric information. While (adverse) selection into networks plays a role, moral hazard (and related punishment strategies) will play a role as well. In particular, the ability to punish others *ex post* and the relative density of the Nyakatoke network may mean that capturing community structure is less relevant. If risk sharing relationships are self-enforcing contracts, simply cutting ties or informing others may be enough to punish someone who reneges (Coate and Ravallion, 1993; Bloch et al., 2008). Furthermore, since this is a repeated game, many rounds of punishment may have already taken place.

3. *Network Sampling*: Network sampling might also be relevant for understanding how the results differ. While I argue that I observe the network ties most salient to the decision, if this is not the case it could be that the communities detected in sampled networks may contain different information than those in census networks. In particular, as fewer nodes are sampled, communities proxy for closer dyadic relationships. I explore this technical explanation in more depth using simulation results (see Section 6.3.1 and Appendix F.1.1).

4. *Cultural and Environmental Factors:* Of course, given that the two datasets are chosen for their network properties, they are quite different in terms of the environmental and cultural context that they are embedded in.

Network measurement and country context non-withstanding, the fact that both the information environment and the form of risk sharing change should not be understated. Despite the multiple differences between the illustrations, these mechanisms are useful in thinking through the external validity of the results from the Colombia experiment. Clearly both information environment and the risk sharing outcome could lead to deviations from the laboratory experiment. However, it is unclear the role each plays and how they might interact.[34]

C.4 The Role of Insurance Groups in Risk Sharing Transfers

The Tanzania data features another unique element: variables indicating co-membership in longstanding insurance groups which provide insurance for funeral expenses and against illness (Dercon et al., 2006). I

[34]For example, were one to observe group formation 'in the wild' with exogenous networks, it may very well have returned a different results than those from the Tanzania illustration, even in spite of the difference in information environment.

estimate several specifications controlling for co-membership in these risk pooling groups, with a specification of interest:

$$\text{Risk Sharing}_{ij} = \alpha + \beta_0 S_{ij} + \sum_{s=1}^{3} \beta_s A^s_{ij} + \gamma C_{ij} + \delta_0 S_{ij} C_{ij} + \sum_{s=1}^{3} \delta_s A^s_{ij} C_{ij} + \eta \text{Group}_{ij} + \varepsilon_{ij}. \tag{11}$$

In particular, that these groups are legible (and indeed, formal) in this context may help understand how well communities proxy for the quasi-formal groups which form for the purpose of risk sharing.

The village data also have a relevant dyadic feature not present in our previous data. In particular, I see if respondents had risk sharing groups in common, which may drive both the formation of networks and also transfers themselves (Fershtman and Persitz, 2021). What role do these groups play in risk sharing transfers in comparison to network structure? I find that there is some overlap in information between these variables. Interestingly, the coefficient on risk sharing is most attenuated by distance-1 connections as opposed to community co-membership. Focusing on specification (5) in Table 7, I find that co-membership in at least one risk sharing group is associated with a 5.1 percentage point increase in the likelihood of making any transfers within a dyad . Notably this correlation is lower, though of similar magnitude of that yielded by community co-membership.

Table 7: The Role of Groups in Effects on Transfers: Tanzania Nyakatoke Network sans Controls

	Any Transfers Within Dyad					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Group(s)	0.114***	0.0627***	0.0528**	0.101***	0.0415*	0.0410*
	(4.23)	(3.56)	(2.92)	(3.97)	(2.39)	(2.41)
Supported		0.714***				0.0853
		(25.29)				(1.49)
Distance-1			0.707***		0.780***	0.706***
			(31.57)		(28.81)	(16.15)
Distance-2					0.139***	0.136***
					(5.38)	(5.28)
Distance-3					0.0300***	0.0291***
					(6.38)	(6.23)
Same Community				0.162***		0.0107
				(6.86)		(0.71)
N	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed from dyadic-robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.5 Addressing Threats to Validity

C.6 Flows on Networks

One limitation of the Tanzania results is that they may not fully appreciate the radius of risk sharing if transfers flow through the network. For example, if a transfer from a distance-2 connection to a distance-1 connection allows for a connection to the origin node, this would not be reflected in these estimates. This is an issue that deserves future attention, though not one I am able to address here.

C.7 Alternative Measures of Transfer Outcomes

To test the robustness of the results, I estimate additional models using alternative transfer outcomes. Using the Tanzania sample, I estimate results for a several outcome definitions, namely if transfers were reciprocal within the dyad and total transfers within the dyad. For reciprocal transfers, results are very similar to transfers overall, except that I observe a slightly smaller radius of risk sharing, with distance-3 nodes entering insignificantly. Considering total transfers, results are again similar, though may be interesting to readers who want to understand the intensity of risk sharing in this network.

C.7.1 Reciprocal Transfers

Reciprocal transfers may be interesting because these dyads are more dependable when need is greatest.[35] For the main specifications, results using reciprocal transfers as the outcome are very similar to the outcomes for any transfers. However, associations are lower in magnitude than those for any transfers owing to the restriction in outcome (see Table 8, specifications 1-4). In contrast to any transfers, I see a smaller radius for reciprocal transfers. In particular, distance-3 connections do not enter significantly here. I also see that distance-2 and 3 connections crowd out co-community membership in explaining transfers, though community transfers do better here than in explaining any transfers (i.e., in terms of t-statistic) (see Table 8, specifications 5-7). Finally, conditional on the differences already reported, the role of groups in reciprocal transfers is close to expected. They help explain reciprocal transfers, feature some difference in information from communities. However, in this case, they are crowded out by distance-2 and 3 connections (see Table 9). Knowing the central role these groups do play in risk sharing, this suggests that whatever their merits I may be somewhat under-powered when analyzing reciprocal transfers and given multicollinearity in regressors.

[35] For example Blumenstock et al. (2016) shows the role of reciprocity driving transfers in response to an earthquake in Rwanda.

Table 8: Effects of Dyadic Relationships on Reciprocal Transfers: Tanzania Nyakatoke Network sans Controls

	Reciprocal Transfers Within Dyad						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.510***			0.0442			0.0457
	(11.67)			(0.60)			(0.62)
Distance-1		0.508***		0.461***	0.522***	0.511***	0.474***
		(13.98)		(8.09)	(13.48)	(13.42)	(8.06)
Distance-2					0.0338***	0.0284**	0.0286**
					(3.39)	(3.06)	(3.09)
Distance-3					0.000222	-0.00146	-0.00141
					(0.07)	(-0.43)	(-0.41)
Same Community			0.113***	0.0282*		0.0193	0.0187
			(6.00)	(2.41)		(1.74)	(1.72)
N	14042	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: The Role of Groups in Effects on Reciprocal Transfers: Tanzania Nyakatoke Network sans Controls

	Reciprocal Transfers Within Dyad					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Group(s)	0.0565***	0.0197*	0.0124	0.0476***	0.00922	0.00863
	(3.94)	(2.32)	(1.45)	(3.66)	(1.07)	(1.02)
Supported		0.508***				0.0455
		(11.54)				(0.62)
Distance-1			0.506***		0.520***	0.473***
			(13.85)		(13.30)	(8.01)
Distance-2					0.0333**	0.0281**
					(3.23)	(2.95)
Distance-3					0.0000595	-0.00155
					(0.02)	(-0.44)
Same Community				0.110***		0.0185
				(5.87)		(1.70)
N	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed from dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.7.2 Total Transfers

Total transfers allow me to go beyond extensive measures of risk sharing to understand the intensity of risk sharing. These results deserve some attention here. I find that distance-1 connections dominate risk sharing when total transfers are used as the outcome. Distance-1 connections are associated with a 3163 TZS increase in the total amount of transfers within the dyad over the five rounds of data collection (Table 10, specifications 1-4). Community co-membership is associated with a 122.9 TZS increase in total transfers, though this is only significant at the 10% level. Again, as distance-2 and distance-3 connections are included, these enter significantly and tend to crowd out detected communities (Table 10, specifications 5-7). Finally, conditional on the differences already reported, groups tend to play a small role in transfer size (Table 11). In fact, in terms of statistical significance groups are only robust to community co-membership. It may be the case that while groups determine risk sharing networks, they do not have additional explanatory power beyond that when it comes to the intensity of private transfers.

Table 10: Effects of Dyadic Relationships on Total Transfers: Tanzania Nyakatoke Network sans Controls

	Total Transfers Within Dyad						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	3046.3***			-123.8			-117.5
	(8.49)			(-0.15)			(-0.15)
Distance-1		3113.3***		3163.0***	3200.6***	3155.7***	3249.4***
		(8.18)		(3.83)	(8.13)	(8.33)	(3.92)
Distance-2					172.3***	150.0***	149.6***
					(4.12)	(3.75)	(3.71)
Distance-3					33.68***	26.74***	26.61***
					(3.58)	(3.39)	(3.42)
Same Community			641.6***	122.9		79.62	81.07
			(4.92)	(1.69)		(1.18)	(1.11)
N	14042	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed from dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: The Role of Groups in Effects on Total Transfers: Tanzania Nyakatoke Network sans Controls

	Total Transfers Within Dyad					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Group(s)	374.5*	154.9	104.2	323.8*	90.34	88.62
	(2.49)	(1.32)	(0.86)	(2.26)	(0.75)	(0.75)
Supported		3025.2***				-119.2
		(8.67)				(-0.15)
Distance-1			3099.3***		3185.6***	3237.0***
			(8.28)		(8.24)	(3.89)
Distance-2					167.1***	145.0***
					(3.94)	(3.45)
Distance-3					32.09***	25.19**
					(3.36)	(2.89)
Same Community				624.8***		79.45
				(4.86)		(1.09)
N	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed from dyadic-robust Standard Errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D Conditional Expectations by Dyadic Relationship

This section details the estimation of these conditional expectations, which are discussed in Section 6.2 and plotted in Figures 2 and 4. There is a great deal of heterogeneity in the dyads of respondents who are co-members in communities, including the distance between dyad members and whether their relationship is supported by a third respondent. Therefore, it may be interesting to examine detected communities in interaction with these other measures. Moreover, this allows me to flexibly estimate excess probability of co-participation in risk sharing conditional on dyad-level features. Extending the model above, I write a full specification which includes interactions between support, friend and family ties, and community co-membership:

$$\text{Risk Sharing}_{ij} = \alpha_v + \beta_0 S_{ij} + \sum_{s=1}^{3} \beta_s A^s_{ij} + \gamma C_{ij} + \delta_0 S_{ij} C_{ij} + \sum_{s=1}^{3} \delta_s A^s_{ij} C_{ij} + \varepsilon_{ijv}. \tag{12}$$

Here, I expect dyads within communities (at a given distance) are more like to match than those dyads between communities. That is, I test $\delta_s > 0$ for $s \in \{0, 1, 2, 3\}$. In addition, excess probability of co-membership can be estimated for each of nine dyad types (relative to dyads who are not connected, supported, or co-community members). For more detail on this exercise. More details of this exercise, including computation for each of these dyad specific means.

Based on specification 12, we specify a conditional expectation (net the constant or fixed effects):

$$E(\text{Risk Sharing}_{ij} | S_{ij}, A^1_{ij}, A^1_{ij}, A^3_{ij}, C_{ij}) - \alpha_v = \beta_0 S_{ij} + \sum_{s=1}^{3} \beta_s A^s_{ij} + \gamma C_{ij} + \delta_0 S_{ij} C_{ij} + \sum_{s=1}^{3} \delta_s A^s_{ij} C_{ij} \tag{13}$$

These expectations can be restated as sums of coefficients from equation 13 and this is done in Table 12. I estimate the empirical analogue of this conditional expectation and present the results in Table 13 (specifically, these estimates are the underlying estimates for Figures 2 and 4).

In contrast to the Colombia results, transfers outcomes depict a much smaller role for communities. Furthermore 4(a)-(c) all show a dramatic drop off as one moves beyond distance-1 connections, despite the ability of distance-2 connections to explain risk sharing.

Table 12: Expectation of Risk Sharing Conditional on Dyadic Relationship

	Community Co-Membership	
Dyadic Relationship	Within Community	Between Communities
Distance-1		
Supported	$\beta_0 + \beta_1 + \gamma + \delta_0 + \delta_1$	$\beta_0 + \beta_1$
Unsupported	$\beta_1 + \gamma + \delta_1$	β_1
Distance-2	$\beta_2 + \gamma + \delta_2$	β_2
Distance-3	$\beta_3 + \gamma + \delta_3$	β_3
Distance-4+	γ	0

Table 13: Estimating Conditional Expectations: Fully Interacted Specifications sans Controls

	Colombia	Tanzania Transfers		
	Same Group	Any	Recip.	Total
	(1)	(2)	(3)	(4)
Supported ($\hat{\beta}_0$)	0.0339	0.0679	-0.00967	1356.7*
	(1.45)	(0.76)	(-0.10)	(2.07)
Distance-1 ($\hat{\beta}_1$)	0.0813***	0.713***	0.465***	1941.4***
	(4.19)	(11.61)	(5.80)	(5.88)
Distance-2 ($\hat{\beta}_2$)	0.0244	0.140***	0.0318***	160.9***
	(1.80)	(5.66)	(3.37)	(4.34)
Distance-3 ($\hat{\beta}_3$)	0.00518	0.0293***	-0.0000	33.06***
	(0.45)	(6.68)	(-0.00)	(3.37)
Same Community ($\hat{\gamma}$)	0.0187	0.0163	0.00271	7.136
	(0.45)	(0.87)	(0.50)	(0.35)
Supported $\times$ Same Community ($\hat{\delta}_0$)	0.0699*	0.0311	0.0870	-3197.4*
	(2.12)	(0.32)	(0.77)	(-2.11)
Distance-1 $\times$ Same Community ($\hat{\delta}_1$)	0.0386	-0.00797	0.0372	2993.2
	(0.82)	(-0.10)	(0.33)	(1.93)
Distance-2 $\times$ Same Community ($\hat{\delta}_2$)	0.0232	-0.0122	0.00443	32.93
	(0.53)	(-0.54)	(0.34)	(0.70)
Distance-3 $\times$ Same Community* ($\hat{\delta}_3$)	-0.00436			
	(-0.10)			
N	88266	14042	14042	14042
Session FE	Yes	No	No	No
SEs	Session†	Dyadic‡	Dyadic‡	Dyadic‡

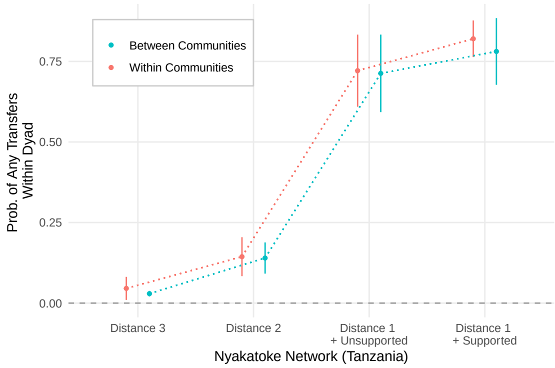
t statistics in parentheses

* Omitted in Tanzania specifications due to perfect multi-collinearity.

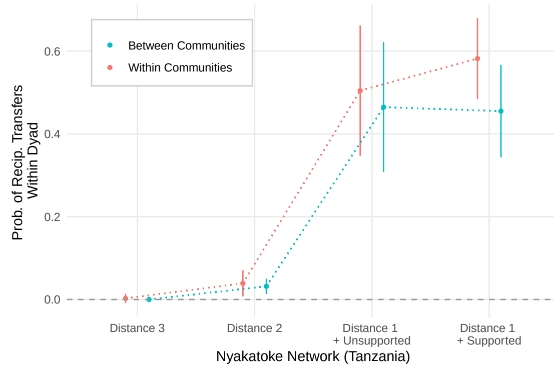
† Indicates cluster robust standard errors clustered at the session level.

‡ Indicates that SEs are dyadic-robust standard errors.

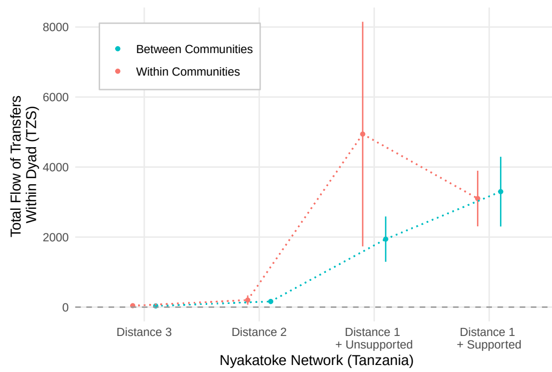
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



(a)



(b)



(c)

Figure 4: The Radii of Risk Sharing in Tanzania: Participation in risk sharing conditional on dyadic relationship featuring estimates from from a fully interacted model. (a) Probability of any transfer within dyad in Tanzania Nyakatoke network. (b) Probability of reciprocal transfer within dyad in Tanzania Nyakatoke network. (c) Total value of transfers within dyad in Tanzania Nyakatoke network.

E Group Specifications: Network Structure and Defaults

E.1 Econometric Specification

Are networks tied to the rate of default within experimental risk pooling groups? To provide context, I build on the analysis from Appendix Table A1 of Attanasio et al. (2012a). Whereas this table presents results using the close friends and family network, I use the friends and family network. In particular, at the group level, I estimate the following specification:

$$\Pr(\text{Default}|G, v) = \alpha_v + \beta N_G + \gamma \text{Density}(G, .) + \delta N_G \times \text{Density}(G, .) + \theta \bar{X}_G + \varepsilon_{Gv} \qquad (14)$$

where N_G is the size of group G, Density$(G, .)$ is one of a set of network densities, $\bar{X}_G$ is a set of controls consisting of group means, and α_v are session fixed effects. As in Attanasio et al. (2012a) I expect $\gamma < 0$ and $\delta > 0$. That is, I expect defaults to fall in network density, but for this effect to attenuate as groups grow larger.

E.2 Variable Construction

I use five definitions of network density: supported density, distance-1 density, distance-2 density, distance-3 density, and co-community density. (By comparison, Attanasio et al. (2012a) computed only distance-1 density within the close friends and family network.) Density is computed by taking in the number of dyads within the group with the given characteristic (e.g., 'are connected') over the total number of dyads within the group. More formally, I compute supported density[36], distance-1 density; and community density:

$$\text{Density}(G, S) = \frac{\sum_{i,j \in G} S_{ij}}{2N_G(N_G - 1)} \tag{15}$$

$$\text{Density}(G, A) = \frac{\sum_{i,j \in G} A_{ij}}{2N_G(N_G - 1)} \tag{16}$$

$$\text{Density}(G, C) = \frac{\sum_{i,j \in G} C_{ij}}{2N_G(N_G - 1)} \tag{17}$$

Distance-s density generalizes network density, includes all dyads of minimum distance less than s:[37]

$$\text{Density}(G, A^s) = \frac{\sum_{i,j \in G} \sum_{t=1}^{s} A_{ij}^t}{2N_G(N_G - 1)}. \tag{18}$$

E.3 Results: Network Structure and Defaults

Network density does not correlate with group level default rates in the friends and family network. The results are both statistically insignificant at all conventional levels of statistical significance, tend to be economically small in magnitude, and are facing in the opposite direction of expectation. For example, a 1 percentage point increase in distance-1 density at the group level corresponds to a 0.036 percentage point *increase* in the default rate (Table 14). On the other hand, when these results are run using the close friends and family network, they appear in the same pattern as Attanasio et al. (2012a), where default falls in network density, and this effect is attenuated as groups grow larger (Table 15). This can be gauged by inspecting Specification (2) in table 15, which is a replication of specification (2) in table A1 of Attanasio et al. (2012a). However, distance-1 density using the close friends and family network is not unique in explaining defaults. In fact, no particular statistic does much better than another, and all are strongly correlated.

In interpreting these results, it is important to note that from the perspective of theory is unclear whether we should see reductions in defaults to be correlated with density. In particular, one could imagine a theoretical model where groups grow only to a size where very few group members default. This size would be endogenous on the underlying network structure. That is, as network structure improves for the purposes of preventing such behavior, group size will grow, testing the limits of such improvements in structure.

[36]Note that I divide the density calculation by two because summing over all entries of the relevant adjacency matrix double-counts the number of connections.

[37]It may more accurately be called shell-s density, though I retain earlier language for rhetorical consistency.

Table 14: Defaults by Group using Colombia Friends and Family Network

	Proportion of Defaults in Risk Pooling Group				
	(1)	(2)	(3)	(4)	(5)
Group Size	0.00350	0.00279	0.00314	0.00524	-0.00222
	(0.93)	(0.66)	(0.71)	(1.00)	(-0.45)
Supported Density	0.0529				
	(0.71)				
Group Size × Supported	-0.0134				
	(-0.71)				
Distance-1 Density		0.0364			
		(0.48)			
Group Size × Distance-1 Density		-0.00769			
		(-0.41)			
Distance-2			0.0130		
			(0.25)		
Group Size × Distance-2 Density			-0.00513		
			(-0.42)		
Distance-3 Density				0.0377	
				(0.71)	
Group Size × Distance-3 Density				-0.00763	
				(-0.65)	
Community Density					-0.0362
					(-0.74)
Group Size × Community Density					0.00930
					(0.86)
N	526	526	526	526	526
Session FE	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Defaults by Group using Colombia Close Friends and Family Network

	Proportion of Defaults in Risk Pooling Group				
	(1)	(2)	(3)	(4)	(5)
Group Size	-0.00221	-0.00441	-0.00586	-0.00656	-0.00415
	(-0.56)	(-1.08)	(-1.23)	(-1.22)	(-0.94)
Supported Density	-0.129*				
	(-2.63)				
Group Size × Supported Density	0.0293				
	(1.90)				
Distance-1 Density		-0.155**			
		(-3.25)			
Group Size × Distance-1 Density		0.0390*			
		(2.61)			
Distance-2 Density			-0.118*		
			(-2.46)		
Group Size × Distance-2 Density			0.0323*		
			(2.61)		
Distance-3 Density				-0.114*	
				(-2.40)	
Group Size × Distance-3 Density				0.0302*	
				(2.42)	
Community Density					-0.120*
					(-2.39)
Group Size × Community Density					0.0282*
					(2.15)
N	526	526	526	526	526
Session FE	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

F Robustness Checks

F.1 Network Measurement

F.1.1 Sampled Networks: Simulation and Results

Since the Tanzania Nyakatoke Network is a village census, I use this data to explore the role of network sampling and resulting measurement error in dyadic relationships. Using this data I randomly sample a proportion of the nodes and keep only the network connections between those households. I then process the data as I would to construct dyadic relationships from the sampled network: support, distance-1, distance-2, distance-3, and community co-membership, and computed using only information from the sampled nodes and their connections. I do two simulations, keeping approximately 25% and 50% of nodes, as dynamics of network sampling might feature some non-linearity (Chandrasekhar and Lewis, 2016). For each simulation, I re-sample the networks 5000 times. As in Smith and Moody (2013), I assume nodes are missing at random.

While nodes might be missing at random, I argue that those who did attend the session should tend to be more central within networks. For example, those municipality members who are more gregarious or socially minded might both have more friends and be more likely to attend events, compared to their more isolated peers. This turns out to be an advantage with regards to network sampling. Smith et al. (2017) stresses the centrality of missing nodes. When lower centrality nodes are removed, closeness is more robust to missing nodes, which translates into more accurate measurement of distance within the network (i.e., distance-1 and 2 connections).

Table 16: Average Correlations Between Dyadic Relationships in Sampled Nyakatoke Network and Full Network

	Census Network				
Sampled Network	Support	Distance-1	Distance-2	Distance-3	Same Comm.
Panel A: 60 Household Samples ($\approx 50\%$)					
Supported	0.83	0.74	-0.16	-0.19	0.22
Distance-1	0.89	1.00	-0.22	-0.26	0.26
Distance-2	-0.13	-0.15	0.69	-0.51	0.16
Distance-3	-0.18	-0.20	-0.20	0.41	-0.05
Same Community	0.38	0.40	0.10	-0.24	0.31
Panel B: 30 Household Samples ($\approx 25\%$)					
Support	0.63	0.57	-0.13	-0.15	0.18
Distance-1	0.89	1.00	-0.22	-0.26	0.26
Distance-2	-0.09	-0.10	0.47	-0.35	0.12
Distance-3	-0.10	-0.11	0.00	0.12	0.02
Same Community	0.50	0.56	0.04	-0.27	0.25

To assess the degree of measurement error, I first check the correlation between dyadic relationships from the sampled network and those from the census network. When 50% of nodes remain, dyadic relationships are strongly correlated with their census counterparts (Table 16). Aside from distance-1 connections, which are not subject to measurement error in this case (Graham, 2020), supported connections feature the highest correlation of 0.83, followed by distance-2 connections (0.69), distance-3 connections (0.41), and co-community membership (0.31). Even when only 25% of nodes remain, dyadic relationships from the sampled networks are correlated to their census counterparts. However, the average correlation falls at different rates. Intuitively, the further the connection is, the more in collapses moving to the 25% sample. For example, while the average correlation for distance-3 connections falls to 26% of its previous value, those for support and community co-membership fall much less (75 and 80% of their previous value, respectively).

Next, I estimate several regressions with (any) transfers as the dependent variable and record the coefficients. I include results from four specifications with the following independent variables: (1) only support, (2) only community co-membership, (3) distance-1, 2 and 3 connections as independent variables, and (4) all of the aforementioned variables. Tables 17) and 18 report the average regression coefficients in 50% and 25% samples, respectively. In general, I see quantitatively small differences between the average regression coefficients and the census estimates. However, a notable counter example is detected community co-membership. Considering single regression, the average coefficient rises as the sample becomes smaller. While the census estimate is 0.167, the average estimate when 50% is sampled is 0.267 and 0.358 when 25% is sampled. My interpretation of this fact is that community detection is picking up more closely connected dyads as fewer nodes are sampled, consistent with correlations found in Table 16. For example, as fewer nodes are sampled detected communities are more correlated with supported nodes and less correlated with distance-2 nodes. In terms of noise, regression coefficients are not particularly noisy in the 50% samples, with SDs tending to be small in percentage point terms. However, as I move to a 25% sample, SDs increase by approximately a factor of two.

Finally, one important point is that in the smaller samples, not all dyadic relationships feature variation, a fact which is reflected in the 'N defined' column of Tables 17) and 18. For example, in 5.7% of cases in the 25% simulation nodes were dropped in such a way that no supported connections existed. In some cases, a coefficient is also undefined because of perfect multicollinearity with other regressors. Where a variable is degenerate, simulations are also omitted in the correlation tables for that variable.

Table 17: Summary of Regression Coefficients (Outcome: Any Transfer) for 60 Household ($\approx 50\%$) Sample of Nyakatoke Network

Coefficient	Model	N Defined	Census Est.	Mean Coef.	St. Dev.
Supported	Support Only	5,000	0.722	0.717	0.051
Distance 1	Distance Only	5,000	0.787	0.760	0.037
Distance 2	Distance Only	5,000	0.141	0.131	0.035
Distance 3	Distance Only	5,000	0.031	0.035	0.017
Same Comm.	Comm. Only	5,000	0.167	0.267	0.063
Supported	Full	5,000	0.086	0.058	0.077
Distance 1	Full	5,000	0.712	0.723	0.054
Distance 2	Full	5,000	0.138	0.129	0.037
Distance 3	Full	5,000	0.030	0.035	0.017
Same Comm.	Full	5,000	0.011	0.005	0.035

Table 18: Summary of Regression Coefficients (Outcome: Any Transfer) for 30 Household ($\approx 25\%$) Sample of Nyakatoke Network

Coefficient	Model	N Defined	Census Est.	Mean Coef.	St. Dev.
Supported	Support Only	4,713	0.722	0.711	0.147
Distance-1	Distance Only	5,000	0.787	0.736	0.077
Distance-2	Distance Only	5,000	0.141	0.113	0.073
Distance-3	Distance Only	4,998	0.031	0.037	0.054
Same Comm.	Comm. Only	5,000	0.167	0.358	0.091
Supported	Full	4,713	0.086	0.050	0.175
Distance-1	Full	5,000	0.712	0.732	0.116
Distance-2	Full	5,000	0.138	0.120	0.088
Distance-3	Full	4,998	0.030	0.039	0.060
Same Comm.	Full	4,990	0.011	−0.019	0.092

F.1.2 Close Friends and Family Network

The network used in Attanasio et al. (2012a) differs in that it is restricted to only close friends and family, where closeness is defined via geographic proximity. Using this network, I re-estimate the main specifications and those related to distance-2 and 3 connections. As can be seen in Table 19, the main specifications are broadly similar. Coefficient estimates are a bit larger and noisier, owing to the sparser nature of the close friends and family network.

Table 19: The Effects Dyadic Relationships on Group Co-Membership: Colombia Close Friends and Family Network sans Controls

	Co-Membership in Risk Pooling Group						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.240***			0.0916**			0.0815
	(8.67)			(2.80)			(1.92)
Distance-1		0.215***		0.0897***	0.191***	0.158***	0.101***
		(9.90)		(4.52)	(8.38)	(8.68)	(3.98)
Distance-2					0.0670***	0.0437***	0.0220
					(5.49)	(3.46)	(1.23)
Distance-3					-0.0233	-0.0356	-0.0320
					(-1.39)	(-1.74)	(-1.62)
Same Community			0.156***	0.0702***		0.0652*	0.0762*
			(8.52)	(4.13)		(2.27)	(2.42)
N	88266	88266	88266	88266	88266	88266	88266
Session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses constructed from session level cluster robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

F.2 Specification Choice

To test robustness to the choice to estimate coefficients using LPM, I first check the predicted outcomes of the LPM, with and without fixed effects. If predictions from LPM lie within the unit interval, the estimated effects should not suffer from the choice of specification (Horrace and Oaxaca, 2006). Second, for those models with predictions outside the unit interval, I re-estimate my models of interest using dyadic logistic regression.

For both illustrations, in specifications I check that do not have fixed effect, predictions lie within the unit interval.

However, in the Colombia illustration, I find that when session fixed effects are employed, a small proportion of predictions lie outside the unit interval, suggesting that LPM may be biased or inconsistent. To assess this in practical terms, I present results from the LPM and logistic regression with and without FE for two of my preferred specifications to compare overall marginal effects. In particular, I replicate specifications (4) and (5) from Table 2.[38]

[38] Unfortunately, I am not able to estimate Chamberlain conditional logit and estimate logit with dummy FEs instead. While one might worry about an incidental parameters problem estimating dummy FEs in a logistic regression with many FE, the fact that the data features a limited number of groups and a large number of dyads should assuage these concerns (in standard panel data this would be a case where T is large relative to n, suggested that bias should be small). Nevertheless, LPM with FEs is preferred given that the non-linearity of the logit model means we cannot obtain estimates that are independent of the group effects, as is possible in a linear model.

Table 20: Logistic Regression: Group Co-Membership on Colombia Friends and Family Network

	Co-Membership in Risk Pooling Group							
	LPM		Logit		LPM		Logit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: LPM Estimates and Logit Overall Marginal Effects†								
Supported	0.0676**	0.0872***	0.0318	0.0567				
	(2.69)	(4.19)						
Distance-1	0.0841***	0.0750***	0.0609	0.0621	0.138***	0.193***	0.0853	0.1596
	(4.11)	(4.92)			(3.76)	(12.00)		
Distance-2					-0.0195	0.0388**	-0.0190	0.0459
					(-0.56)	(3.27)		
Distance-3					-0.0530	0.00661	-0.0632	-0.0019
					(-1.58)	(0.640)		
Same Comm.	0.0306	0.0583***	0.0253	0.0599				
	(1.59)	(5.71)						
Panel B: Odds Ratios								
Supported			0.370*	0.660***				
			(2.56)	(4.28)				
Distance-1			0.709***	0.723***			0.993**	1.858***
			(3.96)	(5.59)			(2.80)	(10.65)
Distance-2							-0.221	0.534**
							(-0.60)	(2.91)
Distance-3							-0.736	-0.0221
							(-1.95)	(-0.12)
Same Comm.			0.294	0.697***				
			(1.48)	(6.51)				
N	88266	88266	88266	86518	88266	88266	88266	86518
Muni. FE	No	Yes	No	Yes	No	Yes	No	Yes

t statistics in parentheses , standard errors clustered at the munipality level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

† For logit estimates, overall marginal effects are computed $\hat{\beta}^{\mathrm{OR}} \times \hat{p}(1-\hat{p})$, where $\hat{p} \approx 0.106$ and $\hat{\beta}^{\mathrm{OR}}$ can be found in Panel B.

Estimates from these models are presented in Table 20. While the logistic results reveal lower total marginal effects than the LPM FE specification, I argue this lays bare the tension between modeling the outcome using a link function and excluding the effect of confounding factors. The logistic regression results without FE are not robust to unobservable factors that might occur at the municipality or session level. We see that OLS estimates fall, without FE, and the overall marginal effects fall yet again in the non FE logit. However, the overall marginal effects from FE logit tend to closely match the LPM FE both qualitatively (i.e. statistical significance in estimates coefficients) but also in magnitude. in these important specifications. Beyond this, LPM with group FE may be better adapted for our particular empirical setting than logit since group formation is rare (Timoneda, 2021).

F.3 Estimation with Selection-on-Observables Approach

Networks are interesting because they are the source of many strategic interactions. When the incentives for network formation rely on many interrelated strategic factors, isolating the causal effect of specific network structure may be difficult. Therefore, rather than arguing that a specific network structure causes co-membership in risk pooling groups, I opt to inform the reader of what assumptions are necessary to credibly interpret the estimates as causal. In particular, even after accounting for the factors detailed below, it could still be the case that the social network structure and risk pooling membership are the result of unobservable differences in dyad level relationships. Thus, a reader would need to believe that I have accounted for the universe of possible factors in order to satisfy conditional unconfoundedness for the following estimates to be taken as causal. To this end, I do account for a battery of potential sources of omitted variable bias from three broad categories: common shocks, popularity, and homophily.

First, in the Colombia illustration I control for common shocks using session fixed effects. Second, certain individuals may be more popular within networks due to their existing characteristics. For example, if it is more prestigious to have rich friends, wealthier people may have more expansive networks than they would otherwise. This effect would manifest itself in both social network structure and choices made in forming experimental risk pooling groups. Third, I also consider other characteristics that might serve as measures of social distance. Respondents who are closer in social, economic, and geographic space tend to be more likely to be connected in social networks (McPherson et al., 2001).

F.3.1 Results from Selection-on-Observables Approach: Colombia

In the Colombia illustration, I include the sum of (log) income, education, risk preferences, and age to control for factors that might drive popularity, and control for the differences in gender, (log) income, education, whether the respondents live in an urban area, risk preferences, and age for social distance. This is consistent with the approach suggested by Fafchamps and Gubert (2007).

Results from selection-on-observables regressions in Colombia broadly accord with their counterparts. These results can be seen in specifications (1)-(4) of Table 21. For main results, patterns of significance (and rough magnitudes) replicate exactly from Table 2. Examining longer walks, there is not a clear pattern of changes in coefficients. However, these regressions do add slightly to the precision of the estimates.

Table 21: Effects of Dyadic Relationships on Group Co-Membership: Colombia Friends and Family Network with Controls

	Co-Membership in Risk Pooling Group						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.198***			0.0903***			0.0963***
	(10.17)			(4.31)			(4.72)
Distance-1		0.176***		0.0751***	0.194***	0.155***	0.0887***
		(11.07)		(4.90)	(12.39)	(9.58)	(4.95)
Distance-2					0.0395**	0.0206	0.0255
					(3.36)	(1.52)	(1.85)
Distance-3					0.00809	0.00169	0.00413
					(0.79)	(0.15)	(0.38)
Same Community			0.112***	0.0547***		0.0487***	0.0442***
			(8.61)	(5.25)		(3.89)	(3.67)
N	88266	88266	88266	88266	88266	88266	88266

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

F.3.2 Results from Selection-on-Observables Approach: Tanzania

In the Tanzania illustration, I include the sum and absolute differences of age and wealth. Additionally, I use measures of social distance between households including if both of the household heads are male, or if one is male and the other female, both household heads have education above a primary level, or if only one household head does, if both households are Muslim (households are either Muslim, Catholic or Lutheran), or if only one household is Muslim, if they belong to the same tribe, and if they belong to the same clan. Similar to in the case of Colombia, this is consistent with the guidance from Fafchamps and Gubert (2007).

I replicate regressions with all three transfer outcomes with these controls. I focus here on Tables 22 and 23, which replicate Tables 6, and 7 from the main text. I find very close accordance between results in these two sets of tables both in terms of the magnitude of estimates and patterns of significance, though estimates fall a bit with the inclusion of controls. Given that the magnitude of estimates of network structure that extend beyond the network neighborhood is smaller, this has a more noticeable effect on distance-2 and 3 connections and community co-membership than distance-1 connections. Reciprocal transfers estimates are summarized in Tables 24 and 25; total transfers in Tables 26 and 27. While these results are also accord closely to the estimates in above, one exception does stick out. In particular, when considering total transfers, distance-2 and 3 connections lose significance when controls are included. This may indicate the importance of the role of homophily in the size of risk sharing transfers beyond the network neighborhood. Indeed, dyadic measures of religion, clan and wealth enter significantly here.

Table 22: Effects of Dyadic Relationships on Transfers: Tanzania Nyakatoke Network with Controls

	Any Transfers Within Dyad						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.667***			0.0553			0.0634
	(23.50)			(0.98)			(1.13)
Distance-1		0.667***		0.609***	0.727***	0.723***	0.673***
		(28.99)		(13.36)	(31.96)	(30.42)	(15.55)
Distance-2					0.108***	0.105***	0.106***
					(5.45)	(5.22)	(5.23)
Distance-3					0.0189*	0.0181*	0.0183*
					(2.56)	(2.52)	(2.55)
Same Community			0.140***	0.0356*		0.00814	0.00735
			(7.06)	(2.42)		(0.57)	(0.51)
N	14042	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 23: The Role of Groups in Effects on Transfers: Tanzania Nyakatoke Network with Controls

	Any Transfers Within Dyad					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Group(s)	0.0877***	0.0476**	0.0381*	0.0789**	0.0320	0.0318
	(3.53)	(2.67)	(2.16)	(3.28)	(1.84)	(1.85)
Supported		0.662***				0.0632
		(23.15)				(1.13)
Distance-1			0.663***		0.723***	0.669***
			(28.53)		(31.12)	(15.39)
Distance-2					0.107***	0.105***
					(5.31)	(5.10)
Distance-3					0.0187*	0.0183*
					(2.48)	(2.46)
Same Community				0.136***		0.00685
				(6.89)		(0.48)
N	14042	14042	14042	14042	14042	14042
Controls	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 24: Effects of Dyadic Relationships on Reciprocal Transfers: Tanzania Nyakatoke Network with Controls

	Reciprocal Transfers Within Dyad						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.490***			0.0374			0.0393
	(11.87)			(0.52)			(0.55)
Distance-1		0.491***		0.452***	0.501***	0.492***	0.461***
		(14.55)		(8.05)	(14.42)	(14.09)	(7.97)
Distance-2					0.0244**	0.0198*	0.0201**
					(3.15)	(2.57)	(2.65)
Distance-3					-0.00183	-0.00331	-0.00319
Same Community			0.0994***	0.0228*		0.0165	0.0160
			(6.30)	(2.19)		(1.58)	(1.57)
					(-0.44)	(-0.77)	(-0.75)
N	14042	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 25: The Role of Groups in Effects on Reciprocal Transfers: Tanzania Nyakatoke Network with Controls

	Reciprocal Transfers Within Dyad					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Group(s)	0.0450***	0.0154	0.00834	0.0387***	0.00662	0.00623
	(3.62)	(1.80)	(1.00)	(3.32)	(0.78)	(0.75)
Supported		0.488***				0.0393
		(11.74)				(0.55)
Distance-1			0.490***		0.501***	0.460***
			(14.40)		(14.24)	(7.94)
Distance-2					0.0242**	0.0199*
					(3.04)	(2.56)
Distance-3					-0.00185	-0.00320
					(-0.44)	(-0.74)
Same Community				0.0978***		0.0159
				(6.14)		(1.55)
N	14042	14042	14042	14042	14042	14042
Controls	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 26: Effects of Dyadic Relationships on Total Transfers: Tanzania Nyakatoke Network with Controls

	Total Transfers Within Dyad						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	2857.0***			-206.6			-202.3
	(8.87)			(-0.26)			(-0.25)
Distance-1		2956.0***		3089.4***	2988.4***	2951.6***	3111.6***
		(8.68)		(3.83)	(8.77)	(8.99)	(3.93)
Distance-2					65.86	47.42	45.94
					(1.83)	(1.16)	(1.05)
Distance-3					1.959	-3.954	-4.562
					(0.07)	(-0.15)	(-0.16)
Same Community			543.8***	83.50		65.96	68.48
			(5.15)	(1.30)		(1.04)	(0.99)
N	14042	14042	14042	14042	14042	14042	14042

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 27: The Role of Groups in Effects on Total Transfers: Tanzania Nyakatoke Network with Controls

	Total Transfers Within Dyad					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Group(s)	294.9*	122.7	74.35	260.7*	70.20	68.87
	(2.30)	(1.15)	(0.68)	(2.10)	(0.64)	(0.63)
Supported		2842.9***				-202.7
		(9.01)				(-0.25)
Distance-1			2947.4***		2979.0***	3103.3***
			(8.76)		(8.86)	(3.90)
Distance-2					63.56	43.98
					(1.74)	(0.99)
Distance-3					1.729	-4.689
					(0.06)	(-0.17)
Same Community				533.1***		67.39
				(5.04)		(0.97)
N	14042	14042	14042	14042	14042	14042
Controls	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses constructed using dyadic-robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$