# Social Network Structure and the Radius of Risk Sharing

Daniel S. Putman\*

This Version: January 27, 2025 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 connections? If so, what is its radius? I examine the radius of risk sharing as a function of 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 sharing groups. Using data from a village census in Tanzania, I find that distance-2 and 3 connections explain risk sharing transfers, but detected community co-membership does not. I address a crucial issue of network sampling in the Colombia illustration using simulation methods, finding my preferred specification is robust to this concern. 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

<sup>\*</sup>Postdoctoral Researcher, University of Pennsylvania Center for Social Norms and Behavioral Dynamics (putmands@gmail.com). This paper has benefited immensely from conversations with Travis Lybbert, Ashish Shenoy, Laura Meinzen-Dick, Arman Rezaee, and Cathy Putman, research assistance from Luke Gooding, as well as comments from several editors and anonymous reviewers. All errors herein are my own.

## 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 sharing arrangements built on trust and reciprocity have 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. 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 (Genicot and Ray, 2003; Fitzsimons et al., 2018; Jain, 2020).

In this paper, I investigate the *radius of risk sharing*. Like the radius of a circle measures the distance from the center of the circle to the perimeter, the radius of risk sharing measures distance from a node to perimeter of the relevant risk sharing network. Where early theory and empirical work placed the village as the radius of risk sharing (Townsend, 1994), more recent empirical work pushed social relationships, as measured by network surveys, to the fore (Fafchamps and Lund, 2003). Still, there are many reasons why risk sharing might extend beyond one's direct friends and family. Theoretical work in economics has pointed to the importance of flows over networks (Bramoullé and Kranton, 2007; Bourlès et al., 2017) and group formation (Genicot and Ray, 2003; Ambrus et al., 2014). Likewise, when sociologists probe people's motivations for who they ask for favors, they often find that these decisions were made intuitively as opposed to deliberatively (Small and Sukhu, 2016).<sup>1</sup> This suggests that the more deliberative approach which tends to be used in network surveys might miss more distant connections which matter for risk sharing. Not as much is known empirically about the radius of risk sharing—most of the work which posits a radius greater than immediate friends and family is either theoretical or simulation based. Moreover, a meaningful distance exists between one's immediate friends and family and the 'six degrees of separation' that might serve as the radius of a village (Strogatz and Watts, 1998; Henderson and Alam, 2022).

What is the radius of risk sharing in social networks? To understand who is near and distant in networks, I examine what network structure best explains co-participation in risk sharing. I draw on *community detection* (Pons and Latapy, 2005; Fortunato, 2010; Newman, 2012) and dyadic regression, an econometric model of network formation (Graham, 2020). I argue that detected communities—clusters of individuals in networks who are closely connected within community and sparsely connected between communities—may serve as a principled method to determine the radius of risk sharing in social networks

<sup>&</sup>lt;sup>1</sup>From Small and Sukhu (2016): in deliberative mobilization "people determine their needs, assess who in their network has the needed attributes, and then turn to that helper." Whereas in intuitive mobilization they circumvent this process, acting spontaneously or incidentally. For example, people were asked for a favor because they happened to be there, as opposed to their being the most useful or one's closest friend.

in addition to established measures of network distance.<sup>2</sup>

To answer this question, I use secondary data from two studies of risk sharing. The first of these is a behavioral risk sharing experiment in Colombia (Attanasio et al., 2012b). The experiment gives participants skin in the game, presenting them with real money gambles and also the ability to share the risk of these gambles with others. Additionally, it includes detailed social network data from participants including friendship and family relationships. Second, I utilize an observational dataset of risk sharing networks and transfers from a census of a village in Tanzania (De Weerdt, 2018).<sup>3</sup> I estimate econometric models of network formation to test whether network structure can explain risk sharing behavior. I use dyadic regression, an approach which treats the dyad-any pair of nodes within the network (connected or unconnected)—as the unit of observation. I characterize network structure around these relationships in the network. More specifically, network structure that might engender risk sharing includes direct connections in social networks (Fafchamps and Lund, 2003), connections a step removed (i.e., 'friends of friends') (De Weerdt and Dercon, 2006), 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). 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 the network structure. Intuitively, these same areas within the network might provide good environments for risk sharing.

Estimates from the risk sharing experiment in Colombia indicate that co-membership in detected communities consistently helps explain co-membership in experimental risk sharing groups, even when controlling for other aspects of network structure. In my preferred specification, co-membership in such a community is associated with a 4.9 percentage point greater propensity to join an experimental risk sharing group together. 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 sharing 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 also estimate a saturated specification which includes all interactions of the network structure variables. I do so to obtain a conditional probability of group co-membership at various distances, which I refer to as radii of risk sharing. If the radius of risk sharing is the distance at which there is *any* co-participation in risk sharing, these radii generalizes this definition to capture distances at which there is a specific *level* of risk sharing. From these, I find that those in the same community are more likely to join the same risk sharing group conditional on every other measure of distance, though sometimes the estimates lose significance in the smaller samples. Most strikingly, I see supported relationships that also lie within communities have a 24 percentage point excess probability of joining the same risk sharing group, relative

<sup>&</sup>lt;sup>2</sup>While community detection algorithms have been widely used in fields where network data is employed, economics has been slow to adopt even as the economics of networks has grown as a field. Two notable exceptions are the use of community detection to study labor mobility (Schmutte, 2014) and to help model exposures in derivative networks (Zema, 2023).

<sup>&</sup>lt;sup>3</sup>This data has been well explored, and forms the basis for results in several related papers including De Weerdt and Dercon (2006), Dercon (2006), Comola and Fafchamps (2014), Comola and Fafchamps (2017), and Henderson and Alam (2022). While this network differs in a number of ways from the Colombia data, as is discussed throughout, it is the closest network for which a census sample is available.

to an 8 percentage point excess probability among a similarly supported pair who are not co-members of a community.<sup>4</sup> This suggests that communities can not only help to understand the outer radius of risk sharing but also may help in identifying 'inner circles' of risk sharing as well.

Not all of these results extend to the Tanzania illustration. While in simpler specifications, detected community co-membership is positively and significantly associated with risk sharing transfers, this effect disappears in my preferred specification, when Distance-2 and 3 connections are accounted for. In this specification co-membership in a community is associated with a 1.1 percentage point greater propensity to make a risk sharing transfer, which is not significant at standard significance levels. However, these results are consistent with those from Colombia in that the radius of risk sharing extends beyond direct connections. In place of the community, I find an association between distance-2 and distance-3 connections with having made a transfer. In particular, distance-2 and 3 connections are associated with 13.8 and 3.0 percentage point increases in the probability of risk sharing transfers, respectively. As might be expected, in this context communities do not identify different radii of risk sharing as they did in the Colombia experiment. Instead, network distance drives the results.

This difference in results does not invalidate the Colombia illustration, nor the Tanzania illustration. The two empirical contexts differ in a number of ways which might explain the discrepancy in results: in addition to cultural and environmental factors more generally, these factors include the nature of the explanatory network, the outcome, sampling, and experimental set-up. Each illustration features its own advantages. The advantages of the Colombia data center around the experimental design and the ability to account for local heterogeneity, whereas the advantages for the Tanzania data center around sampling and *ecological validity*—the degree to which results of the illustrations can be applied to real world settings. Based on these factors, I prefer the Colombia illustration on balance.

The first advantage of the experimental data is that as all risk sharing takes place through the experimental risk sharing groups, the experiment rules out flows of transfers as a consideration. In contrast, in the Tanzania illustration I am not able to account for flows of transfers, the potential for which has been suggested by past work (De Weerdt and Dercon, 2006; Henderson and Alam, 2022). This could in turn reduce the radius of risk sharing found (a downward bias on the coefficients for distance-*s* connections and community co-members). Second, the experimental data have a clear causal ordering which does not admit reverse causality. This is not true of the Tanzania illustration. Third, the experiment draws on data from sessions in 70 different municipalities. Using such data is an advantage because it gives multiple 'draws' from the distribution of potential networks. This is compared to only one village covered by the Tanzania data, which may or may not be typical when considering risk sharing networks, even among similar villages. Moreover, it allows me to use session fixed effects to control for unobservable factors that are correlated with geography.

A fourth advantage is that the experiment allows for considerable coordination before risk sharing groups are formed, but allows participants to default on their group without being observed. This means that networks serve only to limit *adverse selection*, or asymmetric information about the type of people you are sharing risk with, but not *moral hazard*, asymmetric information about the actions they take.<sup>5</sup>

<sup>&</sup>lt;sup>4</sup>This is expressed as *excess* probability of matching as it is in excess of session level fixed effects.

<sup>&</sup>lt;sup>5</sup>That is, the networks help in gathering information about what type of people one is joining a group with (i.e., *adverse* 

This allows me to study the role of networks in risk sharing outside of considerations of monitoring. Based on these features of the experiment, I argue that participants gather information about others in the network before deciding to join the same risk sharing group: they are largely successful in avoiding defaults, despite participant default being unobservable (Attanasio et al., 2012a).<sup>6</sup> While I cannot isolate the exact mechanisms, this process of gathering information may happen through introduction, (explicit or implicit) endorsement, and the ability to talk and interact with other participants during the experiment.

This paper makes three contributions to the literature, one of which derives from a unique robustness check discussed below. First, this paper contributes to the literature on risk sharing among the poor. In particular, the exercise helps to pin down-within this context-a potential radius of risk sharing. This adds value to a literature where models utilize radii of risk sharing which range anywhere from only supported ties (Jackson et al., 2012) to the entirety of a village or administrative unit (Townsend, 1994). Work that empirically explores radii of risk sharing beyond adjacent connections is scant, and is worth documenting here. Dercon et al. (2006) and Lemay-Boucher (2012) document funeral insurance groups-formal, or sometimes quasi-formal groups that extend beyond ones immediate friends and family. De Weerdt and Dercon (2006), within their analysis, show that distance-2 connections matter for non-food consumption smoothing in Tanzania. Finally, Henderson and Alam (2022) study network structure in the same village network in Tanzania, documenting that this network may be unusually good for sharing risk via intermediaries.<sup>7</sup> This work expands on those results. The fact that community co-membership and distance-2 connections matter for experimental risk sharing groups and distance-2 and 3 connections matter for transfers in Tanzania implies a radius of risk sharing that extends beyond those adjacent in risk sharing networks. However, the (sometimes) failure of distance-3 nodes to explain risk sharing participation as well as the erosion of explanatory power over greater distances suggests that risk pools at a relatively more micro level than the village. Thus, one can think of a true radius of risk sharing in this setting (i.e., context, information environment, etc.) as occurring at a meso-level. These results differ from De Weerdt and Dercon (2006) as they they do not test for the role of communities or the role of distance-3 connections in their analysis. Empirically documenting this meso-level radius of risk sharing should direct researchers towards theoretical approaches that model risk sharing at the sub-village level in groups and/or networks (e.g., Genicot and Ray, 2003; Bloch et al., 2008; Ambrus et al., 2014). In contrast, these results do not support models which treat components (connected subgraphs) as the radius of risk sharing (e.g., Bramoullé and Kranton, 2007).8

Second, it contributes to the broader social science literature on favor exchange in social networks. In particular, I provide evidence consistent with the insight that people mobilize (i.e., ask favors of) their social networks intuitively as opposed to deliberatively (Small and Sukhu, 2016). Such deliberative mobilization

selection), but do not help in gathering information about the actions they take-particularly defaulting on commitments (i.e., *moral hazard*).

<sup>&</sup>lt;sup>6</sup>I argue they gather information and not simply trust because of the low incidence of defaults. While these portions of the network may also be home to higher trust e.g., generated by social collateral, social collateral has no bite as defaults are not reported to other group members.

<sup>&</sup>lt;sup>7</sup>Other related empirical papers often study the formation of risk sharing networks, or do not study network structure that extend beyond immediate friends and family.

<sup>&</sup>lt;sup>8</sup>The results here are also qualitatively a different exercise than documenting the rate of risk sharing at various radii (e.g., Townsend, 1994; Kinnan, 2021), as I document the extensive margin, as opposed to the intensive margin conditional on a radius.

would include determining their needs and assessing their network before acting, as one is often asked to do when responding to network surveys. This helps explain the extension of risk sharing beyond direct friends and family, a fact that might be at first counter-intuitive. Very often, improvements in measurement simply make clear to the researcher what the respondents or participants of the study already understand. For example, when we measure risk sharing networks (as opposed to studying villages), we elicit what respondents already know about their networks. However, measurement using community detection and distance-*s* connections often pairs people who are not aware that they lie within each other's risk sharing groups until the moment at which they must form them. In this way, study participants may not fully appreciate the extent of their own risk sharing network.

I address several threats to validity. Most notably, networks in the Colombia illustration are sampled, meaning that some nodes and their associated relationships are unobserved. Network sampling has been shown to lead to measurement error in network statistics and bias in regressions with network statistics (Smith and Moody, 2013; Chandrasekhar and Lewis, 2016). However, to my knowledge this is the first study to address measurement error from node sampling in a dyadic regression framework.<sup>9</sup> Specifically, sampling could result in measurement error in the distance between participants in the friends and family network, if their relationship is supported and if they are in the same detected community. To address this threat to validity, I use a novel network sampling simulation. In these simulations, I keep only a random subset of nodes from the network and their associated links. I then reconstruct relationships between nodes in this sampled network and re-estimate my dyadic regressions of interest.

While I show that sampling does induce measurement error, dyadic relationships computed from sampled data are both strongly correlated with their census counterparts and tend to produce similar dyadic regression estimates. The exception to this is detected communities, where coefficients vary with sampling in 'unconditional' dyadic regressions without controls for other measures of network structure. These results suggest that community detection uncovers closer relationships as fewer nodes are sampled. This, in turn, increases the magnitude of the association between detected communities and risk sharing in unconditional regressions. However, these same coefficients are stable in the full 'longer walks' model—my preferred specification—where other forms of structure are controlled for (i.e., with distance-*s* connections and support). These results suggest that my preferred specification is robust, on average, to measurement error from network sampling.

The results are robust to a number of other empirical exercises. First, I repeat the Colombia analysis using a close friends and family network (as is used in Attanasio et al., 2012a), which restricts friends or family to those dyads living in geographically proximate dwellings. Second, I repeat the Tanzania analysis with alternative measures of risk sharing transfers to demonstrate those results are not sensitive to choices in outcome construction.<sup>10</sup> Given that I am using secondary data, I employ these two exercises to demonstrate that the results are not sensitive to the alternative choices I made in construction of the networks and outcomes. Third, I include a battery of measures of affinity and differences between indi-

<sup>&</sup>lt;sup>9</sup>There is also considerable work on cases where nodes are observed but some or all edges are unobserved (e.g., Breza et al., 2020).

<sup>&</sup>lt;sup>10</sup>When using reciprocal transfers, I do find that the size of associations fall and and distance-3 connections lose significance as might be expected with a more restrictive outcome. However, the pattern is otherwise quite similar.

viduals 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. For the Colombia illustration these include income, education, risk preferences, age, as well as gender controls, and whether the respondents live in an urban area, and in the Tanzania illustration, these include wealth, education, age, gender, religion, and clan. Fourth, I check the robustness of my preferred specifications. I am estimating linear probability models (LPM) in order to include session fixed effects. Therefore, 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.

These results are policy relevant. First, they are most obviously relevant for the financial health enjoyed by those who participate in risk sharing networks. 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 share risk with, all else held equal, the lower the variability of the income received-stabilizing the incomes of all involved in the arrangement. It is increasingly common to ask about risk sharing in measures of financial health (e.g., Karlan and Brune, 2017). Second, access to new financial technology may have the unintended consequence of eroding—or complementing-informal financial and economic relationships (e.g., Dupas et al., 2019; Dizon et al., 2019; Banerjee et al., 2023). These results indicate that the costs-or the benefits-related to network change might be greater than previously thought. Third, understanding the radius of risk sharing can also improve evaluation and design by improving our understanding of the radius of transactions. Transfers are often considered as a component of policy design, particularly transactions which share the gains of treatment (e.g., Janzen et al., 2018). If such spillovers are part of a targeting strategy (i.e., the person is poor, and those they transfer to are also poor), understanding their radius may be of interest to those designing, as a wider radius might erode the ability to target aid.

## 2 Background

### 2.1 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 some 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.<sup>11</sup> In contrast, studies of risk sharing arrangements emphasize the social and economic relationships that serve to mediate risk sharing

<sup>&</sup>lt;sup>11</sup>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.

*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.<sup>12</sup> However, the radius and scale at which informal arrangements can work to share risk is still unclear.

The simplest approach to the radius of risk sharing 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 (e.g., Fafchamps and Lund, 2003; De Weerdt and Dercon, 2006; Jack and Suri, 2014; Blumenstock et al., 2016). However, the risk pool could extend beyond direct friends and family in several ways. In some theoretical models, all members of a network component, a connected subgraph of the network, share risk completely (Bramoullé and Kranton, 2007). In others, sharing is differentiated 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).

In this vein, one could also use group memberships to explore risk sharing connections beyond direct friends and family. Genicot and Ray (2003) explore the formation of such risk sharing 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.<sup>13</sup> 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 of information more than others, which in turn makes punishment of reneging more effective. Finally, Ambrus et al. (2014) build a theoretical model of the effect of 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. Therefore, risk sharing across islands is limited whereas risk sharing within islands is complete.

### 2.2 Community Structure in Risk Sharing Networks

How do we make these larger structures legible within risk sharing networks? One approach would be to use labeled (and therefore formal or quasi-formal) groups.<sup>14</sup> While these groups 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. Generalizing 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

<sup>&</sup>lt;sup>12</sup>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; Kinnan, 2021), moral hazard (Delpierre et al., 2016; Jain, 2020; Kinnan, 2021), transaction costs (Jack and Suri, 2014), and limited commitment (Coate and Ravallion, 1993; Ligon, 1998; Kinnan, 2021). All of these serve to place constraints on risk sharing.

<sup>&</sup>lt;sup>13</sup>Notably these are exogenous networks for which stability is checked; this work does not explain the formation of the networks themselves.

<sup>&</sup>lt;sup>14</sup>For example, risk is sometimes shared explicitly in associations such as funeral insurance groups (Dercon et al., 2006).

related to risk sharing (e.g., Murgai et al., 2002). However, this leaves numerous questions unresolved: For example, should one treat cliques including a greater number of people differently 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 to 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 algorithms seek to detect communities, or 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 direct friends and family. 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. They also tend to accord with economic theory. For example, Bloch et al. (2008) identify dense subgraphs as being stable under various regimes for punishing those who renege.<sup>15</sup> 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. The most common approach is to seek to maximize the *modularity* of detected communities (Fortunato, 2010). The intuition behind modularity is that it is maximized when there are dense connections within communities and sparse connections between communities. Newman (2012) defines modularity as follows: Let  $A_{ij}$  be the ijth entry of the adjacency matrix and  $C_{ij}$  the ijth entry of the community co-membership matrix. Let  $d_i$  and  $d_j$  be the degrees of nodes i and j, respectively, defined as the number of connections they have in the network. Let m be the number of edges in the graph. Modularity is expressed:

$$Q = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \left( A_{ij} - \frac{d_i d_j}{2m} \right) C_{ij}$$
(1)

To understand this statistic, consider a random graph as a counterfactual, where all pairs of nodes have an equal probability of linking. In such a graph, the probability of a link between nodes i and j 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. I then take the difference between actual links  $A_{ij}$  and expected links  $d_i d_j / 2m$  and weight this by if nodes are in the same community. Finally, to express modularity, aggregate to the graph level and normalize by twice the number of links present.

Directly maximizing modularity is a difficult problem in large networks, so approaches instead depend

<sup>&</sup>lt;sup>15</sup>In particular, they are stable under strong punishment (where the reneger is forced into autarky) and intermediate punishments (where people within some radius of the reneged upon cut ties with the reneger). This includes 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.)



(a) Distance and Communities

(b) Support and Communities

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.

on approximation algorithms whose differences from direct maximization are bounded (Fortunato, 2010). My approach to uncovering these latent structures is based on one of these approximation algorithms, and uses random walks through the network. In this setting, a random walk moves from node-to-node in the network by way of edges, randomly selecting the next node it visits among those in that nodes' network. 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. In this approximation algorithm, modularity serves as a validation tool to choose between possible community assignments. When all links occur within the communities, this statistic will be at its highest, reflecting a modular community structure.

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 in a component have been merged into one community, recording each potential community assignment along the way.
- 5. This process yields a *dendrogram*, or a hierarchical diagram documenting the potential community assignments from the algorithm and merges. I compare the modularity of all potential community assignments, and choose the highest modularity.

Figure 1 presents the detected communities in a single session network. Figure 2(a) depicts a dendrogram from the same single session network, while Figure 2(b) depicts the modularity of each community assignment. Appendix A.1 presents the formula for distance computations. Appendix A.2 explores the importance of the length of walks. Additionally, in Appendix A.3 I visualize the potential community assignment associated with each cut of the dendrogram.

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, it 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 (as is analyzed in Bloch et al., 2008). 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 (see Appendix A.4 for a specific discussion of the edge-betweenness algorithm).

## 3 Data and Context

#### 3.1 The Risk Sharing 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).<sup>16</sup> 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 sharing 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 sharing 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 sharing

<sup>&</sup>lt;sup>16</sup>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 F2.



Figure 2: *Walktrap* Community Detection: (a) Dendrogram produced by *Walktrap* community detection for the example session network above. This plot visualizes the merges of nodes into communities according to the *Walktrap* algorithm, with each merge providing a potential community assignment. The colored boxes represented the observed community detection where the dendrogram is cut. 'Node ID' refers to the unique identifier for each node. (b) I cut the dendrogram using a statistic called modularity. Modularity is maximized at cut 14, so I cut the dendrogram for the observed community assignment. The community detection here is the same visualized in Figure 1

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 sharing groups.

The first round of the risk sharing 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 them. As can be seen in Table B1, 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 sharing groups in which winnings would be pooled and shared equally, which would be declared before the second set of meetings took place.<sup>17</sup> 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

<sup>&</sup>lt;sup>17</sup>Participants were given around an hour to an hour and a half (during lunch) to form their groups.

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 5.4.1.

#### 3.2 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 an 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, 2002). In addition, individuals are asked to list those who they could personally rely on for assistance, which forms the risk (cross-sectional) sharing network of interest.

#### 3.3 Variable Construction

Nodes represent participants, and edges represent their direct relationships (e.g., if they are friends or family).<sup>18</sup> For a given network g, let  $N = \{1, ..., n\}$  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  $A_{ij} = \mathbf{1}(ij \in E)$ .<sup>19</sup> In this representation, the adjacency matrix will be defined symmetrically:  $A_{ij} = A_{ji}$  for all  $i, j \in N$ . This means that I will treat networks as undirected.

#### 3.3.1 Colombia Risk Sharing Experiment

**Risk Sharing Groups** The outcome of interest is whether or not a dyad of individuals joined the same experimental risk sharing group (even if they later reneged). Being in a risk sharing group with the other

$$\mathbf{1}(\text{condition}) = \begin{cases} 1, & \text{if condition is true} \\ 0, & \text{if condition is false} \end{cases}$$
(2)

<sup>&</sup>lt;sup>18</sup>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.

<sup>&</sup>lt;sup>19</sup>Here, and throughout the text 1(.) denotes the indicator function,

member of the dyad is referred to as co-membership in the risk sharing group. Note that groups are nonoverlapping: each participant can only be in one group. Formally, for  $i \in \text{Group}_i$  and  $j \in \text{Group}_j$ , I define  $\text{Group}_{ij} = \mathbf{1}(\text{Group}_i = \text{Group}_j)$ .

**Social Networks** The explanatory variables of interest are constructed from the network survey 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.<sup>20</sup> In contrast to Attanasio et al. (2012a), which used only geographically proximate connections, the network is unrestricted by location.

#### 3.3.2 The Nyakatoke Network Data

**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 define an indicator variable for if any transfers are made within the dyad in any round. I also construct two alternative outcomes based on the transfers, presented in Appendix B.2.

**Risk Sharing Network** For the Tanzania data, as opposed to friends and family, respondents are asked "Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour?" The network differs from social networks in that it is a direct elicitation of the risk sharing network. Comola and Fafchamps (2014) finds that this network is best understood as the result of household *desire-to-link*. A connection in this desire-to-link network means one of the two respondents might want to ask for assistance in the future.<sup>21</sup> I define the network to be undirected and unilateral as to create greater consistency with the Colombia illustration.

#### 3.3.3 Dyadic Relationships

I start by forming an undirected and unweighted graph of the explanatory network, g. For Colombia, this is the friends and family network, while in Tanzania, this is the risk sharing network. I say i and j are connected (or  $ij \in g$ ), if either i recognizes j in the explanatory network 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

<sup>&</sup>lt;sup>20</sup>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).

<sup>&</sup>lt;sup>21</sup>This is as opposed to unilateral or bilateral network formation, which require one or both members of the dyad to accede to the link.

connections as  $A_{ij}^2 = \mathbf{1}(\min \operatorname{distance}(i, j) = 2)$  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  $A_{ij}^3 = \mathbf{1}(\min \operatorname{distance}(i, j) = 3)$ . Figure 1(a) plots distances from an origin node within a network for a single session in the Colombia illustration. Finally, supported relationships are any dyad where there is a link in the explanatory network where both dyad members share a common connection. Formally, I define supported connections as  $S_{ij} = \mathbf{1}(ij \in g \text{ and } \exists k \text{ such that } ik, jk \in g)$ . Figure 1(b) plots the supported connections of an origin node within a network for a single session in the Colombia data

#### 3.3.4 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 the Colombia data and Figure B1 plot communities in the Tanzania Nyakatoke data.

#### 3.4 Summary Statistics

#### 3.4.1 Outcomes

Participation varied 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, for a total of 2378 participants across 70 session. 86.9% of participants chose to join a risk sharing 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 all dyads are within an experimental risk sharing group. In the Tanzania illustration, about 14.6% of dyads featured some kind of transfers, meaning each respondent household had made transfers with around 8.6 other households. Table B2 summarizes these outcomes.

#### 3.4.2 Explanatory Networks

To describe overall network structure, details of the social network characteristics for the different social networks are presented in Table B3. A single session network from Colombia is visualized in Figure 1. The average respondent names about 3.5 friends or family members who attend the session. This yields an average density of 0.056, meaning that of all potential connections (within the session), about 1 in 20 exist. The village network from Tanzania is presented in Figure B1. 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. Based on this, we know that both networks are dense, in the sense that the rate of connection is sufficient so that most of the nodes are in a single component (Achlioptas et al., 2009). As documented in Appendix B.3 these networks are also clustered with short average path length.

#### 3.4.3 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, as I don't have a compelling reason to change this parameter, I use the default of length four.<sup>22</sup> For more on how walk length matters for detected communities, see Appendix A.2.

Results of community detection are presented alongside network characteristics in Table B3. In the Colombia networks, I find communities of average size 4.5 and modularity of 0.44. While it is clear that distance-*s* connections extend to a radius beyond direct friends and families, this larger scale of communities *vis a vis* the risk sharing network suggests a larger radius as well. If we were to take detected communities as the radius of risk sharing, they would provide more potential risk sharing partners than would be available via direct connections (3.5), increasing the scale of risk sharing. Detected communities are larger in the Tanzania illustration, reflecting more interconnected network data. These communities can be seen in Figure B1. I find communities of size 10.8 and modularity of 0.30. Though modularity is not directly comparable across different network sizes, it is likely that lower modularity is partially explained by lower density and clustering in this network. Communities provide an average additional 1.6 potential risk sharing partners relative to direct connections (9.8 as compared to 8.2). 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.

In Appendix B.4, I include a comparative analysis of community composition, identifying what factors impact homophily across contexts and what factors lead to popularity (i.e., who finds themselves in larger communities). This analysis is consistent with the idea that communities in the Colombia illustration represent closer relationships than in the Tanzania illustration.

#### 3.4.4 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. A similar number of dyads are within communities in the two networks: 19.2% of dyads are within communities in Tanzania, in comparison to 18.0% of dyads in Colombia. In both networks, dyads within communities have closer relationships than those between communities, both in terms of features like network distance and support. Nevertheless, detected communities would represent a wider radius of risk sharing than direct connections. In Colombia, a plurality (45.2%) of dyads within communities are distance-2 connections. Additionally, some dyads within communities were distance-3,

<sup>&</sup>lt;sup>22</sup>In addition to removing researcher degrees of freedom which might lead to 'cherry picking' of results, this ensures consistency across empirical illustrations to ease comparison.

	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	

Table 1: Characterizing Dyadic Relationships by Community Co-Membership

*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 the nodes in the dyad lie in separate communities. Supported dyads are directly connected with at least one additional node connected to both. Unsupported are connections with no other node connected to both.

and very few (less than 1%) are distance-4 or greater. Likewise, the majority of dyads within communities in Tanzania (56.5%) are distance-2 connections. Comparing between Colombia and Tanzania, the closer connections within communities in Colombia are consistent with the greater density and clustering in these networks.

## 4 Empirical Strategy

To test the explanatory power of various measures of the radius of risk sharing, I use dyadic regression, an econometric model of network formation. This method is easily interpretable and ideal for a first cut at understanding the radius of risk sharing. In these regressions, each pair of participants–whether connected or unconnected within the network–is treated as an observation. My approach explicitly summarizes broader network structure at the dyad level to better take account of the complex dynamics at play in social networks.<sup>23</sup>

## 4.1 Simple Specification

I start with a simple specification that seeks to explain participation in risk sharing using connections in the explanatory network. 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

<sup>&</sup>lt;sup>23</sup>To increase the kinds of network motifs I account for would mean estimating Subgraph Generation Models (Chandrasekhar and Jackson, 2023). While such models would certainly add value, they are less easily interpretable models of network formation than dyadic regression and would add considerable complexity to the analysis at hand.

is defined as a direct connection if i or j recognize friendship or family ties. Second, this relationship is supported if there is an 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 3.3.) The main specification is as follows:

Risk Sharing<sub>*ij*</sub> = 
$$\alpha_v + \beta_0 S_{ij} + \beta_1 A_{ij} + \gamma C_{ij} + \varepsilon_{ijv}$$
 (3)

where Risk Sharing<sub>ij</sub> measures if i and j joined a group together,  $\alpha_v$  is a session fixed effect,  $A_{ij}$  is an indicator equal to 1 if direct connection is present,  $S_{ij}$  is an indicator equal to 1 if i and j have a supported connection, 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. Similarly,  $\gamma > 0$  indicates that detected communities explain risk sharing.

#### 4.2 Longer Walks: Increasing the Radius of Risk Sharing

While detected communities may be one way we see increased radius of risk sharing, 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 about distance in the network. To test 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

Risk Sharing<sub>*ij*</sub> = 
$$\alpha_v + \beta_0 S_{ij} + \sum_{s=1}^3 \beta_s A^s_{ij} + \gamma C_{ij} + \varepsilon_{ijv}$$
 (4)

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  $\gamma$ , tests of  $\beta_s$  might indicate that risk sharing extends beyond direct connections. If rejected, these tests indicate that those further-flung members in an individual's social network are good candidates for sharing 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.,  $\beta_1 > \beta_2 > \beta_3 > 0$ .

### 4.3 Fully Saturated Specification

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:

$$\operatorname{Risk}\operatorname{Sharing}_{ij} = \alpha_v + \beta_0 S_{ij} + \sum_{s=1}^3 \beta_s A_{ij}^s + \gamma C_{ij} + \delta_0 S_{ij} C_{ij} + \sum_{s=1}^3 \delta_s A_{ij}^s C_{ij} + \varepsilon_{ijv}.$$
(5)

Here, I expect dyads within communities at a given distance are more likely to match than those dyads between communities at the same distance. 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).

Based on specification 1, I specify a conditional expectation (net the constant or fixed effects):

$$E(\text{Risk Sharing}_{ij}|S_{ij}, A_{ij}^{1}, A_{ij}^{1}, A_{ij}^{3}, C_{ij}) - \alpha_{v} = \beta_{0}S_{ij} + \sum_{s=1}^{3} \beta_{s}A_{ij}^{s} + \gamma C_{ij} + \delta_{0}S_{ij}C_{ij} + \sum_{s=1}^{3} \delta_{s}A_{ij}^{s}C_{ij}$$
(6)

These expectations can be restated as sums of coefficients from equation (6) (see Table C1).

#### 4.4 Estimation, Standard Errors, and Robustness

In the Colombia illustration, I estimate the above specifications using linear probability models (LPM) in order 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 within 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 D.5 as robustness checks. Based on the structure of the Colombia data, I only include dyads that were in the same session. I correct for non-independence of standard errors by clustering at the session level. For the Tanzania Nyakatoke Network data, I correct for non-independence using dyadic-robust standard errors (Fafchamps and Gubert, 2007; Cameron and Miller, 2014; Tabord-Meehan, 2019).<sup>24</sup> For both illustrations, the standard errors are robust to heteroskedasticity, which is of particular importance when estimating LPMs.

I perform a number of other exercises to assess the validity of results. I address network sampling using a simulation approach. I re-estimate the Colombia illustration with a different social network, made up of geographically proximate friends and family. I re-estimate the Tanzania illustration with alternate transfer outcomes. 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 D.

<sup>&</sup>lt;sup>24</sup>While I might also use dyadic-robust standard errors for Colombia, cluster robust standard errors at a session level tends to be more conservative.

	Co-Membership in Risk Sharing 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 Dyads	88266	88266	88266	88266	88266	88266	88266
Session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2: Associations Between Dyadic Relationships and Group Co-Membership

Notes: Data is the Colombia friends and family network. Results based on 2378 participants in 70 sessions. All regressions feature session level fixed effects, cluster robust standard errors at the session level, and no additional controls. t statistics in parentheses. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## 5 Results and Discussion

### 5.1 Social Network Structure and Experimental Risk Sharing in Colombia

#### 5.1.1 Simplest Specifications: Support, Neighborhood, and Community

How well do these measures of the network explain co-membership in experimental risk sharing groups? Focusing on the first four specifications reported in Table 2, supported friends or family, friends or family, and community co-membership enter positively and significantly (I 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. Among these specifications, I prefer to focus on column (4), which includes all three variables. As I document in the robustness checks, specifications which are more inclusive tend to feature a more stable coefficient for community co-membership, so my preferred specification will be discussed in the next section. Here, being adjacent in the network is associated with a 7.5 percentage point increase in the likelihood of joining the same risk sharing group, and being in the same community is associated with a 5.8 percentage point increase in the likelihood of joining the same risk sharing group, and being in the same risk sharing 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 sharing group if they are close (geographically proximate) friends or family. Second, this reinforces that support drives risk sharing 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 explain risk sharing above and beyond these previously explored network measures.

#### 5.1.2 Longer Walks: Distance-s Connections

While the measures included in the simple specification seem to do well on their own and in concert, the same is not true for 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. I take column (7) as my preferred specification for understanding the association between community co-membership and co-membership in risk sharing groups. First, because the magnitude of the estimate falls as distance-2 and distance-3 connections are added, this is a more conservative estimate of the association. Second, as I show in Section 5.4.1, the coefficient on community co-membership is relatively stable in this specification as nodes are sampled. In this specification, being in the same community is associated with a 4.9 percentage point increase in the likelihood of joining the same risk sharing group.

## 5.2 Risk Sharing Network Structure and Transfers in Tanzania

## 5.2.1 Simplest Specifications: 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 3. While support, distance-1 connections, and community co-membership all 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 adjacent in the network 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 sharing group.

#### 5.2.2 Longer Walks: Distance-s Connections

While 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

	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

Table 3: Associations Between Dyadic Relationships and Dyadic Transfers

*Notes:* Data is the Tanzania Nyakatoke Network. Results based on 120 household respondents. t statistics in parentheses constructed using dyadic-robust standard errors. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

in Table 3, 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 comembership 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.

## 5.3 The Radii of Risk Sharing

A saturated (i.e., fully interacted) regression specification allows us to inspect participation in risk sharing conditional on distance and community membership. In doing so, a picture of strong and weak ties in these networks emerges—multiple radii of risk sharing. Using estimates from the fully interacted specifications, I construct conditional expectations of participation in risk sharing and plot these in Figure 3. The underlying calculations and estimates for these empirical specifications are presented in Tables C1 and C2, respectively and are further discussed in Appendix C.

Turning to the Colombia illustration, visualized in Figure 3(a), I find that the shorter the distance between two respondents, the higher the probability of co-membership regardless of community status. As



Figure 3: The Radii of Risk Sharing in Colombia and Tanzania: Participation in risk sharing conditional on dyadic relationship featuring estimates from from a fully interacted model. (a) Excess probability of comembership in risk sharing group in Colombia friends and family network. (b) Probability of any transfer within dyad in Tanzania Nyakatoke network.

seen in previous regression results, distance-3 ties tell us little about the probability of co-membership, whereas shorter distances are more informative. Not only does the probability of co-membership in a risk sharing group tend to increase for dyads that are closer 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 sharing group formation at every distance.<sup>25</sup>

In contrast to the Colombia results, the Tanzania results depict a much smaller role for communities and network support to help detect strong and weak ties. Figure 3(b) shows a dramatic drop off in the probability of transfers as one moves beyond distance-1 connections, despite the ability of distance-2 and 3 connections to explain risk sharing in this context. This pattern is similar when alternative measures of transfers are used, as can be seen in Figure C1.

### 5.4 Threats to Validity and Limitations

#### 5.4.1 Network Sampling and Measurement Error

The Colombia illustration may face a unique form of measurement error, due to network sampling.<sup>26</sup> If the network observed among session members is not the salient network for the risk sharing 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 sup-

<sup>&</sup>lt;sup>25</sup>Weakly better in the sense that I cannot always reject the null hypothesis that the means are equivalent.

<sup>&</sup>lt;sup>26</sup>While this is conceptually related to Chandrasekhar and Lewis (2016), the implications differ due to the dyadic setting.



Figure 4: Associations Between community co-membership and risk sharing from network sampling simulations. The full model includes support, distance-*s* variables, and community co-membership. 'Comm. only' includes community co-membership as the sole regressor. Horizontal lines (and 0% estimates) indicate the estimate with no network sampling. 95% confidence bands represent variation in coefficients from sampling. (a) Results from 5000 network sampling simulations with the Nyakatoke network data and any transfer as an outcome. (b) Results of 500 network sampling simulations with the Colombia session networks and risk sharing group co-membership as an outcome.

ported 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 were these networks not sampled.

First, I argue that the effect of measurement error on my 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 sharing groups, the broader social network will be incompletely remembered, while the within session network is directly observable.<sup>27</sup> Furthermore, no information can 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

<sup>&</sup>lt;sup>27</sup>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.

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 simulate the process of sampling to better understand the impact of this element on results from the dyadic regressions. The simulation proceeds as follows. In each network, I randomly sample respondent nodes (households/participants) and keep only the network connections between those sampled nodes. I then proceed to generate a dyadic dataset of the remaining nodes and check the correlation between dyadic variables constructed from the subsampled dataset and their counterparts using the full sample available to me. Finally, I re-estimate regressions and compare average estimated coefficients to those estimated via the full samples.

I start with a simulation using the Tanzania Nyakatoke Network, as it forms a census of a single village. 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, regression coefficients tend to be stable for non-community measures of network structure. Third, it guides my interpretation of the estimated coefficients on detected communities towards full models. I present these estimates (as well as a those from sampling 75% of nodes) in Figure 4(a). In particular, the association of detected communities with transfers tends to vary more with sampling than other variables in the unconditional regressions. The estimated associations between detected communities and transfers in the unconditional regression increase in magnitude as fewer nodes are sampled. Detected communities tend to be smaller when detected using smaller networks and therefore proxy for closer relationships as fewer nodes are sampled. This interpretation is borne out by correlations between sampled and census variables. However, in the full model, where close associations are accounted for by the other measures of network structure, associations between communities and transfers are stable (if insignificant).

I replicate these three insights in the Colombia friends and family networks. First, variables computed in subsampled networks are highly correlated with their values in the session level data available to me. Second, the regression coefficients for supported distance-1, 2, and 3 connections are relatively stable with regard to sampling, in both the unconditional and the conditional regressions. Third, the association between communities and risk sharing group co-membership rises somewhat in the unconditional regression as fewer nodes are sampled, but is stable in the longer walks model. Based on the results from the previous simulation exercise, I would expect the coefficient to rise in these regressions as more nodes were excluded. We see exactly this pattern in Figure 4(b). This gives an obvious preference to the coefficient estimate from the longer walks model as compared to the unconditional model since these results do not vary with sampling in either the first or second simulation. See Appendix D.1 for detailed methods and additional results.

#### 5.4.2 Close Friends and Family Network

I test for robustness to network definition in the Colombia illustration, using the close friends and family network which is featured in Attanasio et al. (2012a). In this network, *close* friends and family are friends and family who are also geographically proximate. I find the results are robust to this alternative network,

though coefficients are larger and somewhat noisier. See Appendix D.2 for detailed results.

### 5.4.3 Alternative Measures of Transfers and Flows on Networks

I test for the robustness of the outcome definition in the Tanzania illustration, using two alternative measures of transfers: if transfers were reciprocal and the total transfer value. I find that when reciprocal transfers are used, there is a similar pattern of results in my main specification. However, as this measure is more restrictive, the magnitude of estimates falls on distance-1 and 2 connections, and distance-3 connections are no longer associated with transfers (Table D7). Using total transfer value, we once again find a similar pattern of results, with positive and significant absolute transfer value for distance-2 and 3 connections, though with small magnitudes compared to direct connections (Table D8). Further discussion of these results can be found in Section D.3.

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 another transfer to the origin node to take place, this would not be reflected in these estimates. In particular, Henderson and Alam (2022) suggest that this network is nearly optimally structured for such flows to matter in risk sharing, and indeed, De Weerdt and Dercon (2006) show that distance-2 connections matter for smoothing non-food consumption, evidence that may be suggestive of this point. Outside of the involvement of the intermediary, flows of transfers are perfect substitutes for direct transfers. If flows of transfers are important, for purposes of understanding the radius of risk sharing, flows of transfers would positively bias the coefficient on distance-1 connections, and would negatively bias the coefficients on distance-2 and 3 connections, and would likely negatively bias the coefficient on community co-membership (given that most people in the same community are indirect connections). This is an issue that deserves future attention, though not one I am able to address within the scope of this paper.

#### 5.4.4 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 sharing experiment was conducted after real-world networks were realized, the results should not suffer from the possibility of reverse causality.<sup>28</sup> Second, while common shocks may play a role in both networks and group formation, I am able to control for these in estimates 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. These fixed effects alleviate some concerns with omitted variable bias, particularly those factors that are correlated across municipalities.

<sup>&</sup>lt;sup>28</sup>This is not true of the Tanzania illustration. The outcome in that illustration is transfers while the explanatory network is a desire-to-link risk sharing network. While a such a risk sharing network precedes future transfers in the causal ordering, transfers are retrospective. When queried who one would ask for assistance, they might recall previous transfers. Therefore, past transfers (or non-transfers) may impact one's future desire-to-link.

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 in the main text. Appendix D.4 describes my variable selection approach and presents estimates for both the Colombia and Tanzania illustrations.

#### 5.4.5 LPM and Logistic Regression

I test for robustness to specification choice, particularly 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). This includes all specifications from the Tanzania illustration. 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 fixed 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 fixed effects in my logistic specification, marginal effects are similar to those reported in LPM with fixed effects. Therefore, I stand by my use of LPM with fixed effects 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 D.5.

#### 5.4.6 Ecological Validity

All behavioral experiments will face some limitations due to ecological validity, and the Colombia illustration is no different (Berkowitz and Donnerstein, 1982). With that said, the experiment is well designed to capture risk sharing as if in the real world. The experiment is framed within existing patterns of risk sharing in the context. The experiment takes place close to where respondents live, with other members of their community—likely risk sharing partners. Groups are formed in a familiar social setting. The gambles and risk sharing games are played for real stakes: the safest gamble amounts to an expected value of about 21.2% of average daily household income in the sample; the riskiest gamble 42.4%.

#### 5.5 Discussion

#### 5.5.1 Meso-Level Risk Sharing

Taken together, the results start to tell a story of 'meso-level' risk sharing. If the most macro-level risk sharing occurs at the village level (or the session level) and the most micro-level risk sharing occurs with those adjacent in the network, the results here are at an intermediate level. Risk sharing falls far short of the diameter, or the longest minimum distance between two nodes of the largest components. However, it extends beyond those adjacent in the network. 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, in

the group formation experiment, community co-membership does well in explaining participation in risk sharing among connections of distance-2. This suggests that these community detection tools may be useful in bounding the radius of risk sharing when groups are loosely defined or illegible to outsiders. This broader radius of risk sharing implies that households may be better equipped to smooth consumption than we would otherwise have imagined.

#### 5.5.2 Group Formation, Trust, and Adverse Selection in Colombia

Why should network structure (and community structure in particular) matter in the Colombia risk sharing 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).<sup>29</sup> 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 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. In particular, in data from the World Values Survey, only 16% of men and 12% of women responded that most people can be trusted (Sudarsky, 2018).<sup>30</sup> This low trust environment motivates my interpretation: 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.

Additionally, 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 their group.<sup>31</sup> 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 salient, allowing for participants to coordinate their responses in such a game (Attanasio et al., 2012a).

<sup>&</sup>lt;sup>29</sup>Technically, the model in Ambrus et al. (2014) need not be limited commitment, but the set-up might be implied by such a model.

<sup>&</sup>lt;sup>30</sup>In particular, this is in response to the general trust question, "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"

<sup>&</sup>lt;sup>31</sup>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 is known there are dishonest types in the group, the gains to sharing risk fall for trustworthy types. This pushes out trustworthy types, thereby decreasing the realization at which dishonest types will default, leading to an overall collapse in the group.

#### 5.5.3 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 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 question, several mechanisms may play a role: sampling, the explanatory networks, outcomes, and the role of asymmetric information in that illustration, and cultural and environmental factors more broadly.

- Network Sampling: While the Nyakatoke networks are a village census of households, the Colombia sessions are a sample of individuals. While I argue that in Colombia I observe the network ties most salient to the decision to form groups, 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. I explore this technical explanation in more depth using simulation results (see Section 5.4.1 and Appendix D.1). In particular, as fewer nodes are sampled, communities proxy for closer dyadic relationships. However, coefficients are stable when controlling for other forms of network structure. Even in these specifications, results differ. This suggests we should look elsewhere to explain differences in these two contexts.
- 2. **Outcomes:** Risk sharing outcomes do not match one-to-one across these two illustrations. While in Colombia experimental risk sharing groups are the outcome of interest, in Tanzania I observe dyadic transfers instead. One reason that this might matter is that experimental risk sharing groups are by nature non-overlapping, as are the community detection methods we use. Consider a node on the edge of two risk sharing groups, who must choose only one. Community structure may predict their choice because it captures those nodes they have stronger connections to. This is not the case with dyadic transfers, where one could make transfers with both would-be risk sharing groups. Another reason this may matter is the potential for transfers to flow over networks, something which the Colombia experiment rules out. As is discussed in Section 5.4.3, this may lead to downward bias on the coefficients for indirect relationships.
- 3. **Explanatory Networks:** The networks differ in what they capture. While the Colombia experiment captured social networks (friends and family), the Tanzania data captures an *ex ante* risk sharing network directly.<sup>32</sup> In particular, Comola and Fafchamps (2014) shows that the Nyakatoke network is best explained within a desire-to-link framework, meaning it represents not the risk sharing network *per se*, but instead represents those whom one might ask in the future given the risk sharing network. It might be interesting to see if detected communities based on other social network ties predict such desire-to-link risk sharing networks in this data or in other contexts.
- 4. **Information Environment** 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 informa-

<sup>&</sup>lt;sup>32</sup>I.e., it is based on responses to the question "Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour?"

tion. 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, many rounds of punishment may have already taken place, and may impact the explanatory network we now observe.

5. Cultural and Environmental Factors Of course, given that the two datasets are chosen for their network properties, they are drawn from different populations which are quite different in terms of the environmental and cultural context that they are embedded in. For example, the Nyakatoke network has relatively low clustering compared to what tends to be reported in networks (Henderson and Alam, 2022). This departs from other known favor networks (e.g., Jackson et al., 2012, which documents such structure in Indian villages).

The fact that the outcome, information environment, and explanatory network differ should not be understated. Indeed, the fact that results do not extend to the Tanzania illustration suggest an interesting direction for future work. Intermediate illustrations might provide some sense of what exactly communities capture in the Colombia illustration that helps them explain risk sharing behavior.

#### 5.5.4 Additional Analyses

I include two additional analyses in the Appendix which are not for purposes of robustness, but that add richness to the set of results for those readers who may be interested. First in Appendix E I add insurance group co-co-membership to the slate of predictors in the Tanzania illustration to understand their role in risk sharing transfers. Second, in Appendix F I study the relationship between various measures of network density and defaults in the Colombia illustration.

## 6 Conclusion

Using dyadic regression, I test explanatory power of measures of network structure in explaining experimental risk-sharing outcomes. In doing this, I uncover how the radius of risk sharing 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 sharing 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 method novel in application to risk sharing. In addition, distance-2 connections sometimes explain co-membership in experimental risk sharing groups, though these estimates are not as strong or stable as the other three. In contrast, not all of these results extend to the Tanzania Nyakatoke data. In my preferred specification, distance-2 and 3 connections crowd out co-membership in detected communities and supported connections also lose their explanatory power. Both sets of results point toward risk sharing that takes place at a level between the village (session) level and bilateral level. This finding might guide how we think about the welfare derived from informal risk sharing. 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 be conservative in this regard. When considering the literature on risk sharing, theoretical models that allow for this kind of 'meso-level' risk sharing become more intriguing, such as the work by Genicot and Ray (2003), Bloch et al. (2008), and Ambrus et al. (2014), among others. Still, more work is needed to understand how community detection might provide value across contexts which might vary by the form of risk sharing (groups or bilateral), the type of network used, the information environment, the explanatory network, and more. For example, communities might be used as the relevant group in consumption smoothing risk sharing regressions (like those seen in De Weerdt and Dercon, 2006; Kinnan, 2021) to understand the relevance of this network structure when transfers might flow through networks.

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 those adjacent in the network. In particular, it is often the case that network formation is guided by *homophily*, or the principle 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 direct friends and family? Interesting applications of communities include the study of assortative matching on risk preferences (Putman, 2020). Finally, while risk sharing is an exciting application of community detection, it may prove valuable for other places where social networks are relevant to the provision of goods. Questions still remain for the use of network science to understand risk sharing in networks. Where empirical tools match imperfectly with theory, clarifying the links between empirical and theoretical modeling of these phenomena will be crucial.

## References

- D. Achlioptas, R. M. D'Souza, and J. Spencer. Explosive Percolation in Random Networks. *Science*, 358 (3186):2000–2002, 2009. ISSN 1095-9203. doi: 10.1126/science.1167373.
- 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. *Review of Economic Studies*, 2023. doi: doi:10.1093/restud/rdad065.
- 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.
- L. Berkowitz and E. Donnerstein. External Validity Is More Than Skin Deep. *American Psychologist*, 37(3): 245–257, 1982.
- 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, Nov. 2023.
- 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, World Bank Group, 2018.
- S. Dercon. Risk, Growth and Poverty: What do we know, what do we need to know? Technical Report 148, 2006.
- S. Dercon, J. De Weerdt, T. Bold, and A. Pankhurst. Group-based funeral insurance in Ethiopia and Tanza-

nia. 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.
- 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.
- V. Labatut and J.-M. Balasque. Detection and Interpretation of Communities in Complex Networks: Practical Methods and Application. In A. Abraham and A.-E. Hassanien, editors, *Computational Social Networks*, pages 81–113. Springer London, London, 2012. ISBN 978-1-4471-4047-4 978-1-4471-4048-1. doi: 10.1007/978-1-4471-4048-1\_4.
- P. Lemay-Boucher. Insurance for the Poor: The Case of Informal Insurance Groups in Benin. Journal of Development Studies, 48(9):1258–1273, Sept. 2012. ISSN 0022-0388, 1743-9140. doi: 10.1080/00220388. 2012.693172.
- 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.
- 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.
- 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. Putman. *Social Network Structure and Risk in the Developing World*. PhD thesis, University of California, Davis, 2020.
- I. M. Schmutte. Free to Move? A Network Analytic Approach for Learning the Limits to Job Mobility. *Labour Economics*, 29:49–61, 2014. doi: 10.1016/j.labeco.2014.05.003.
- 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 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, World Values Survey, 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.
- S. M. Zema. Uncovering the network structure of non-centrally cleared derivative markets: Evidence from large regulatory data. *Empirical Economics*, 65(4):1799–1822, Oct. 2023. ISSN 0377-7332, 1435-8921. doi: 10.1007/s00181-023-02396-9.

## **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):

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

 $P_{ik}^s$  is the probability that a walk starting at node *i* ends its 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:

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

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:

$$P_{C,k}^{s} = \frac{1}{|C|} \sum_{i \in C} P_{ik}^{s}.$$
(3)

#### A.2 Length of Random Walks for Walktrap

Performance (as measured by modularity) and community size differ by walk length. For the Colombia data, modularity is relatively static with respect to path length. However, 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 size 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.3 Example of Walktrap Algorithm Merges

I depict the process of the *Walktrap* algorithm in Figures A1, A2, and A3. Each network depicts a potential community assignment, were the dendrogram cut at that step of the algorithm. While this is just one example, in Figure A1 the first merges tend to be within the largest eventual communities and occur in supported relationships. The last few cuts in A3 (15-18) correspond to declines in modularity.



Figure A1: *Walktrap* cuts 2-5: Single session Colombia friends and family network. Node color indicates potential community assignments at each cut of the dendrogram, with the associated modularity printed at top. Nodes marked with their ID.


Figure A2: *Walktrap* cuts 6-9: Single session Colombia friends and family network. Node color indicates potential community assignments at each cut of the dendrogram, with the associated modularity printed at top. Nodes marked with their ID.



Figure A3: *Walktrap* cuts 10-18: Single session Colombia friends and family network. Node color indicates potential community assignments at each cut of the dendrogram, with the associated modularity printed at top. Nodes marked with their ID.

### A.4 Edge Betweenness Community Detection

An alternative approach to community detection is called *edge betweenness*, and utilizes the algorithm from Girvan and Newman (2004). This algorithm uses the counts of shortest paths on edges in the network between nodes to split up the network. What does this mean, exactly? To compute edge betweenness we find the shortest path in the network between every pair of nodes. Using these paths, we count the number of paths that lie on a given edge in the network. If there are multiple shortest paths of equal length, partial credit is awarded, 1/2 for two paths, 1/3 for three, etc. The intuition is that these edges would carry a high load in a risk sharing arrangement. Conversely, information could be controlled across these edges,



(a) Walktrap Algorithm

(b) Edge Betweenness



making monitoring or learning about others' reputation hard (for this reason betweenness is sometimes called "brokerage" centrality).

The algorithm proceeds as follows:

- 1. Compute edge betweenness for all edges
- 2. Find the edge with the highest betweenness, and remove it from the network
- 3. Recalculate betweenness for all edges that remain
- 4. Repeat until all edges have been removed

This leaves us with a set of potential community assignments based on network components (connected subgraphs). Every time the network is split into multiple components, this is a potential community assignment. As in the Walktrap algorithm, these are compared using modularity, and the assignment with the highest modularity is selected.

The community assignments produced by edge betweenness are similar to those from Walktrap. Community co-membership between the two methods is highly correlated (Pearson r = 0.72) and both are about equally correlated with co-membership in an experimental risk sharing group. For example, see Figure A4 which plots community detection in one session. We can see that assignments are the same except one community which is split in two. This specific session is more highly correlated than average, but is shown for consistency with the main text (Pearson r = 0.92). Edge betweenness features similar average modularity: averaging a score of 0.43 as compared to 0.44 from Walktrap.

### **B** Data Appendix

### **B.1** The Risk Sharing 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

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

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

$$Var(\bar{y}) = \frac{1}{N_G} \sum_{\ell=1}^{6} q_\ell^2 \times Var(y_\ell)$$
<sup>(2)</sup>

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

Payoff					
	Gamble	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

Table B1: Incentive Structure for the Gamble Choice Game

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

#### **B.2** Alternative Measures of 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, the main outcome reported in the analysis from the main text. Second, I define an indicator variable for if these transfers were reciprocal. That is, if 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

### Table B2: Summary of Dyadic Outcomes

Colombia	Obs	Mean/Prop	Std. Dev.	Min	Max
Risk Sharing 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

are made.

### **B.3** Network Summary

More details of overall network structure can be found in Table B3. Here I will summarize a number of network statistics, including density, clustering, and closeness.

- 1. 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, 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.
- 2. The clustering coefficient measures the transitivity of social connection. This is computed by dividing the number of closed triplets of nodes by the number of all triplets, open or closed, where a triplet is a connected set of three nodes.<sup>33</sup> 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.
- 3. 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.<sup>34</sup> In the Nyakatoke network, closeness 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).

<sup>&</sup>lt;sup>33</sup>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?

<sup>&</sup>lt;sup>34</sup>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 each other, on average. As closeness approaches zero, nodes are very far, or more likely in separate components.



Figure B1: Nyakatoke Risk Sharing Network with Communities Overlaid. Detected communities are represented by both shaded regions and node color.

	С	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
C	(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)	-

## Table B3: Characteristics of Networks and Detected Communities

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.

### **B.4** Interpretation and Comparison of Communities

It is interesting and important to understand what detected communities represent in the two empirical contexts. Approaches for doing so tend to explore either the network structure of communities or the attributes of nodes or relationships in these communities (Labatut and Balasque, 2012). I explore the network structure between and with communities elsewhere but I will briefly summarize those findings, before focusing on the attributes of nodes.

Table B3 presents the average number and size of communities in both contexts. We see that while there are similar numbers of communities per network, the size of these communities is much larger in the Tanzania Nyakatoke Network. Specifically, the average community is 10.8 nodes as a opposed to 4.5 in the Colombia Friends and Family network. Consistent with these findings, Table 1 shows that relationships found within communities are closer within communities as opposed to between communities, and that relationships within the Colombia Friends and Family Network are closer *vis a vis* the Tanzania Nyakatoke Network.

Turning to node attributes, I analyze homophily and popularity in the communities, and for reference, the explanatory networks. While other approaches exist, in keeping with the methods presented in the main text, I will deploy dyadic regression. To make the analysis comparable across contexts, I am limited to three sociodemographic variables found in both contexts: age, education level, and consumption. It could of course be the case that variables omitted here (clan, religion, etc.) drive matching in these contexts.

I regress the network of interest on functions of these covariates which measure how socially close or distant the members of the dyad are. In these regressions I include an indicator for whether respondents have the same level of education (and whether this is high or low level of low education), the absolute difference and sum of age, and the absolute difference and sum of log consumption. For Tanzania, I have only the level of education completed as opposed to years of education, therefore I generate level of education for Colombia. Since the end of primary school is a key stopping point for education in both context, I collapse education to two levels, depending on whether respondents finished primary school. To make consumption comparable, I convert to \$PPP/month. Then I take the log of consumption and the absolute difference and sum of these values in the dyad.

The results for Colombia and Tanzania are presented in Tables B4 and B5, respectively. There is overall little evidence of homophily or differences in popularity across age, education and consumption at least when considering significant differences in matching associated with these factors. This may be a question of noise as much as signal, effects need to be quite large compared to the outcome mean to gain significance when using robust standard error estimates.

Table B4 documents homophily in age in both the Colombia friends and family network as well as the community network. A twenty-five year difference in age is associated with a 2.3 percentage point reduction in the probability of being in the same detected community. We also see some homophily by education, at least among high education respondents. Both members of the dyad being high education is associated with a 1.9 percentage point increase in the probability of being in the same detected community. While insignificant, there is also a negative association between the community network and the absolute differences in log consumption. Table B5 documents a significant associated with a 2.0 percentage point increase in the sum of age is associated with a 2.0 percentage point increase in the probability of matching. It is consistent with the notion that communities measure closer relationships in Colombia (and in sampled data more generally) that evidence of homophily is found there and not as consistently in the Tanzania data. In other places, while not significant, the results are sometimes consistent across illustrations. In particular, the coefficient on the sum of age is and both having a high education level are both positive for all four networks.

	(1)	(2)
	Friends and Family Network	Community
Age Sum	0.000216	0.000530
	(1.13)	(1.47)
Age Diff.	-0.00122***	-0.000903**
	(-5.34)	(-2.71)
Both High Ed.	0.00251	$0.0188^{*}$
	(0.38)	(2.02)
Both Low Ed.	0.00979	0.0109
	(1.75)	(0.92)
Sum Log Cons.	-0.00344	-0.00448
	(-0.59)	(-0.42)
Diff. Log Cons.	0.00327	0.00191
	(0.66)	(0.21)
Constant	0.134	0.192
	(1.81)	(1.43)
Outcome Mean	0.103	0.180
N	88266	88266

Table B4: Assortative Matching in Colombia Social Networks and Communities

*t* statistics from cluster robust standard errors in parentheses

	(1)	(2)
	Risk Sharing Network	Community
Age Sum	$0.000800^{*}$	0.000910
	(2.54)	(1.04)
Age Diff	-0.000603	0.00101
	(-1.55)	(1.46)
Both High Ed.	0.00704	0.0456
	(0.79)	(1.10)
Both Low Ed.	-0.00375	-0.00283
	(-0.27)	(-0.08)
Sum Log Cons.	0.0208	-0.0116
-	(1.60)	(-0.35)
Diff. Log Cons.	-0.0151	0.0207
C	(-1.33)	(0.60)
Constant	-0.110	0.147
	(-1.21)	(0.68)
Outcome Mean	0.0697	0.192
N	11556	11556

Table B5: Assortative Matching in Tanzania Risk Sharing Networks and Communities

t statistics from dyadic robust standard errors in parentheses

# C Conditional Expectations by Dyadic Relationship

This section details the estimation of these conditional expectations, which are discussed in Section 5.3 and plotted in Figures 3 and C1. I estimate the empirical analogue of this conditional expectation and present the results in Table C2 (specifically, these estimates are the underlying estimates for Figures 3 and C1). In contrast to the Colombia results, transfers outcomes depict a much smaller role for communities. Furthermore C1(a)-(b) show a dramatic drop off as one moves beyond distance-1 connections, despite the ability of distance-2 connections to explain risk sharing.

	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	$\beta_1 + \gamma + \delta_1$	$eta_1$				
Distance-2	$\beta_2 + \gamma + \delta_2$	$eta_2$				
Distance-3	$\beta_3 + \gamma + \delta_3$	$eta_3$				
Distance-4+	$\gamma$	0				

Table C1: Expectation of Risk Sharing Conditional on Dyadic Relationship



Figure C1: 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 reciprocal transfer within dyad in Tanzania Nyakatoke network. (b) Total value of transfers within dyad in Tanzania Nyakatoke network.

	Colombia	Ta	nzania Trans	fers
	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 $ imes$ Same Community ( $\hat{\delta}_0$ )	0.0699*	0.0311	0.0870	-3197.4*
	(2.12)	(0.32)	(0.77)	(-2.11)
Distance-1 $ imes$ Same Community $(\hat{\delta}_1)$	0.0386	-0.00797	0.0372	2993.2
	(0.82)	(-0.10)	(0.33)	(1.93)
Distance-2 $ imes$ Same Community $(\hat{\delta}_2)$	0.0232	-0.0122	0.00443	32.93
	(0.53)	(-0.54)	(0.34)	(0.70)
Distance-3 × Same Community <sup><i>a</i></sup> ( $\hat{\delta}_3$ )	-0.00436			
	(-0.10)			
N	88266	14042	14042	14042
Session FE	Yes	No	No	No
SEs	Session <sup>b</sup>	$Dyadic^c$	$Dyadic^c$	$\operatorname{Dyadic}^{c}$

Table C2: Estimating Conditional Expectations: Fully Interacted Specifications sans Controls

t statistics in parentheses

 $^{a}$  Omitted in Tanzania specifications due to perfect multicollinearity.

<sup>b</sup> Indicates cluster robust standard errors clustered at the session level.

 $^{c}$  Indicates that SEs are dyadic-robust standard errors.

# **D** Robustness Checks

### D.1 Sampled Networks: Simulation and Results

### D.1.1 Tanzania Nyakatoke Network

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).

	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 House	hold Sample	s ( $\approx 25\%$ )			
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

Table D1: Average Correlations Between Dyadic Relationships in Sampled Nyakatoke Network and Full Network

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 D1). 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 the correlation 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 D2 and D3 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 D1. For example, as fewer nodes are sampled detected communities are more correlated with supported nodes and less correlated with distance-2 nodes. Additionally, it is worth discussing the fact that the average coefficient on community detection is negative in the 25% sample when controls are included. This seems to be related more closely to the fact that communities do not explain transfers more than anything else, as the estimate is only 0.03 different that the census estimate, and would not be significantly different than zero. 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 D2 and D3. For example, in 5.7% of cases in the 25% simulation, nodes were selected 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.

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 D2: 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	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

Table D3: Summary of Regression Coefficients (Outcome: Any Transfer) for 30 Household ( $\approx 25\%$ ) Sample of Nyakatoke Network

#### D.1.2 Colombia Session Networks

I repeat the analysis above for the Colombia session networks, this time further sampling these already sampled networks. Again, 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 D4). Aside from distance-1 connections, supported connections feature the highest correlation of approximately 1, followed by distance-2 connections (0.82), co-community membership (0.59), and distance-3 connections (0.48), and . Even when only 25% of nodes remain, dyadic relationships from the sampled networks are correlated to their census counterparts.

Table D4: Average Correlations Between Dyadic Relationships in Colombia Session Networks and Sampled Subnetwork

	Session Network				
Sampled Network	Support	Distance-1	Distance-2	Distance-3	Same Comm.
Panel A: 50% Samj	oles				
Support	1.00	0.86	0.43	0.28	0.47
Distance-1	0.86	1.00	0.32	0.30	0.51
Distance-2	-0.16	-0.19	0.82	0.45	0.23
Distance-3	-0.15	-0.17	-0.35	0.48	-0.11
Same Community	0.55	0.63	0.45	0.35	0.59
Panel B: 25% Samp	oles				
Support	1.00	0.86	0.43	0.27	0.47
Distance-1	0.86	1.00	0.32	0.29	0.50
Distance-2	-0.16	-0.19	0.82	0.45	0.23
Distance-3	-0.15	-0.17	-0.35	0.48	-0.11
Same Community	0.65	0.76	0.41	0.32	0.52

Next, I estimate several regressions with risk sharing group co-membership as the dependent variable and record the coefficients. As before, I include results from four specifications with the following indepen-

dent 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. Table D5 reports the average regression coefficients in 50% and 25% samples. Again, for support and distance-*s* connections, I see quantitatively small differences between the average regression coefficients and the session level estimates. In the univariate regression, as before, the coefficient on same community rises slightly with sampling.<sup>35</sup> While the census estimate is 0.115, the average estimate when 50% are sampled is 0.132 and 0.152 when 25% of nodes are sampled. This follows the same logic as above, that community detection is picking up more closely connected dyads as fewer nodes are sampled. However, when considering the full model, the coefficient on same community only changes very slightly, from 0.049 to approximately 0.048 in the 50% sample and the 25% sample.

			Coefficient Estimates			
			50% Sı	ıbsample	25% Sı	ıbsample
Coefficient	Model	Session	Mean	St. Dev.	Mean	St. Dev.
Support	Support	0.197	0.199	0.016	0.202	0.034
Distance-1	Distance	0.193	0.195	0.015	0.197	0.031
Distance-2	Distance	0.039	0.040	0.009	0.040	0.020
Distance-3	Distance	0.007	0.008	0.007	0.007	0.017
Same Community	Comm.	0.115	0.132	0.014	0.152	0.028
Support	Full	0.092	0.100	0.025	0.108	0.057
Distance-1	Full	0.087	0.082	0.023	0.076	0.056
Distance-2	Full	0.023	0.030	0.010	0.035	0.020
Distance-3	Full	0.002	0.007	0.007	0.008	0.017
Same Community	Full	0.049	0.048	0.015	0.048	0.036

Table D5: Summary of Regression Coefficients for Samples of Colombia Session Networks

<sup>&</sup>lt;sup>35</sup>This seems to be smaller relative to the presented coefficient size, but this is probably due to the fact that we are sampling after sampling has already taken place, meaning we are not adjusting the degree of sampling as much.

## D.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 D6, 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.

		Co-Membership in Risk Sharing 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		

Table D6: The Effects Dyadic Relationships on Group Co-Membership: Colombia Close Friends and Family Network sans Controls

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

### D.3 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.

### **D.3.1** Reciprocal Transfers

Reciprocal transfers may be interesting because these dyads are more dependable when need is greatest.<sup>36</sup> 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 D7, 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 D7, 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 E2). 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 given multicollinearity in regressors.

			Reciprocal	Transfers	Within Dva	d	
	(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

Table D7: Effects of Dyadic Relationships on Reciprocal Transfers: Tanzania Nyakatoke Network sans Controls

t statistics in parentheses constructed using dyadic-robust standard errors

<sup>&</sup>lt;sup>36</sup>For example Blumenstock et al. (2016) show the role of reciprocity driving transfers in response to an earthquake in Rwanda.

### **D.3.2** 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 D8, 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 D8, specifications 5-7). Finally, conditional on the differences already reported, groups tend to play a small role in transfer size (Table E3). 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.

			Total Ti	ransfers Wit	hin 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

Table D8: Effects of Dyadic Relationships on Total Transfers: Tanzania Nyakatoke Network sans Controls

t statistics in parentheses constructed from dyadic-robust standard errors

### D.4 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 sharing 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 sharing 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 sharing 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).

### D.4.1 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 D9. 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.

#### D.4.2 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 D10 and E4, which replicate Tables 3, and E1 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 those directly adjacent 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 D11 and E5; total transfers in Tables D12 and E6. While these results are also accord closely to the estimates in above, one exception does stick out. In particular, when considering

		Co-Membership in Risk Sharing 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		

Table D9: Effects of Dyadic Relationships on Group Co-Membership: Colombia Friends and Family Network with Controls

*t* statistics in parentheses

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

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 those adjacent in the network. Indeed, dyadic measures of religion, clan and wealth enter significantly here.

		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			

Table D10: Effects of Dyadic Relationships on Transfers: Tanzania Nyakatoke Network with Controls

*t* statistics in parentheses constructed using dyadic-robust standard errors

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

Table D11: Effects of Dyadic Relationships on Reciprocal Transfers: Tanzania Nyakatoke Network with Controls

			<b>D</b> · · · · ·			1				
		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

	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		

Table D12: Effects of Dyadic Relationships on Total Transfers: Tanzania Nyakatoke Network with Controls

t statistics in parentheses constructed using dyadic-robust standard errors

### **D.5** 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 fixed effects for two of my preferred specifications to compare overall marginal effects. In particular, I replicate specifications (4) and (7) from Table 2.<sup>37</sup>

Estimates from these models are presented in Table D13. While the logistic results reveal lower total marginal effects than the LPM fixed effects 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 fixed effects are not robust to unobservable factors that might occur at the municipality or session level. We see that OLS estimates fall, without fixed effects, and the overall marginal effects fall yet again in the non fixed effects logit. However, the overall marginal effects from fixed effects logit tend to closely match the LPM fixed effects both qualitatively (i.e., statistical significance and sign) but also in magnitude. Beyond this, LPM with group fixed effects may be better adapted for this particular empirical setting than logit since group co-membership is a rare outcome (Timoneda, 2021). Specifically, in my preferred specification (Table D13 column 8) detected community co-membership enters significantly with a 5.3 percentage point increase in group co-membership associated with being in the same detected community.

 $<sup>^{37}</sup>$ Unfortunately, I am not able to estimate Chamberlain conditional logit and instead 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 (this is akin to the case in standard panel where T is large relative to n, suggesting 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.

	Co-Membership in Risk Sharing Group								
	LF	PM	Lo	git	LI	PM	Lo	git	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Panel A: LPM	Estimates and	l Logit Overa	ll Marginal	$\operatorname{Effects}^a$					
Supported	0.0676**	$0.0872^{***}$	0.0351	0.0625	$0.0648^{*}$	0.0920***	0.0323	0.0674	
	(2.69)	(4.19)			(2.55)	(4.52)			
Distance-1	0.0841***	0.0750***	0.0672	0.0685	0.0541	0.0866***	0.0342	0.0836	
	(4.11)	(4.92)			(1.52)	(4.81)			
Distance-2					-0.0349	0.0227	-0.0379	0.0308	
					(-1.00)	(1.63)			
Distance-3					-0.0572	0.00203	-0.0747	-0.0083	
					(-1.70)	(0.18)			
Same Comm.	0.0306	0.0583***	0.0279	0.0661	0.0459***	0.0488***	0.0445	0.0532	
Den al D. O I Ia	(1.59)	(5./1)			(3.67)	(4.09)			
Panel B: Odds	Ratios		0.070*	0 ( ( 0 * * *			0.041*	0 11 1 * * *	
Supported			$0.370^{\circ}$	0.660			0.341	0.711	
			(2.56)	(4.28)			(2.32)	(4.65)	
Distance-1			0.709***	0.723***			0.361	$0.882^{***}$	
			(3.96)	(5.59)			(1.02)	(4.23)	
Distance-2							-0.400	0.325	
							(-1.08)	(1.48)	
Distance-3							-0.788*	-0.0881	
							(-2.08)	(-0.47)	
Same Comm.			0.294	0.697***			0.470***	0.561***	
			(1.48)	(6.51)			(4.08)	(4.01)	
N	88266	88266	88266	86518	88266	88266	88266	86518	
Session FE	No	Yes	No	Yes	No	Yes	No	Yes	

Table D13:	Logistic 1	Regression:	Group	Co-Memb	ership o	n Colombia	Friends an	nd Family	Network
	~	0							

t statistics in parentheses \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001<sup>a</sup> Overall marginal effects are computed  $\hat{\beta}^{\text{OR}} \times \hat{p}(1-\hat{p})$ , where  $\hat{p} \approx 0.106$  and  $\hat{\beta}^{\text{OR}}$  can be found in Panel B.

# E The Role of Insurance Groups in Risk Sharing Transfers

### E.1 Summary

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.

### E.2 Insurance Groups

I also have risk sharing 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<sub>i</sub> be the set such that Group<sub>i</sub>  $\in$  Groups<sub>i</sub>, Any Group<sub>ij</sub> =  $\mathbf{1}(\exists \text{Group } \in \text{Groups}_i)$ 

## E.3 Empirical Strategy and Results

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

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

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 the 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 E1, 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.

		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)				
Ν	14042	14042	14042	14042	14042	14042				

Table E1: The Role of Groups in Effects on Transfers: Tanzania Nyakatoke Network sans Controls

t statistics in parentheses constructed from dyadic-robust Standard Errors

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

\_

		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				

Table E2: The Role of Groups in Effects on Reciprocal Transfers: Tanzania Nyakatoke Network sans Controls

\_

t statistics in parentheses constructed from dyadic-robust standard errors

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

=

		1	otal Transfe	ers 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)
Ν	14042	14042	14042	14042	14042	14042

Table E3: The Role of Groups in Effects on Total Transfers: Tanzania Nyakatoke Network sans Controls

t statistics in parentheses constructed from dyadic-robust Standard Errors

	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		

Table E4: The Role of Groups in Effects on Transfers: Tanzania Nyakatoke Network with Controls

t statistics in parentheses constructed using dyadic-robust standard errors

	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		

Table E5: The Role of Groups in Effects on Reciprocal Transfers: Tanzania Nyakatoke Network with Controls

t statistics in parentheses constructed using dyadic-robust standard errors

	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		

Table E6: The Role of Groups in Effects on Total Transfers: Tanzania Nyakatoke Network with Controls

t statistics in parentheses constructed using dyadic-robust standard errors

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

=

## F Group Specifications: Network Structure and Defaults

### F.1 Econometric Specification

Are networks tied to the rate of default within experimental risk sharing 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:

$$\Pr(\text{Default}|G, v) = \alpha_v + \beta N_G + \gamma \text{Density}(G, .) + \delta N_G \times \text{Density}(G, .) + \theta X_G + \varepsilon_{Gv}$$
(1)

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  $\alpha_v$  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.

#### F.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<sup>38</sup>, distance-1 density; and community density:

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

$$Density(G, A) = \frac{\sum_{i,j \in G} A_{ij}}{2N_G(N_G - 1)}$$
(3)

$$Density(G,C) = \frac{\sum_{i,j\in G} C_{ij}}{2N_G(N_G - 1)}$$
(4)

Distance-s density generalizes network density, includes all dyads of minimum distance less than s:<sup>39</sup>

Density
$$(G, A^s) = \frac{\sum_{i,j \in G} \sum_{t=1}^s A_{ij}^t}{2N_G(N_G - 1)}.$$
 (5)

### F.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 F1). 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 F2). This can be gauged by inspecting Specification (2) in table F2, which is a replication of specification (2) in table A1 of Attanasio

<sup>&</sup>lt;sup>38</sup>Note that I divide the density calculation by two because summing over all entries of the relevant adjacency matrix doublecounts the number of connections.

<sup>&</sup>lt;sup>39</sup>It may more accurately be called shell-*s* density, though I retain earlier language for rhetorical consistency.

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.

	Proportion of Defaults in Risk Sharing 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 $\times$ Supported	-0.0134					
	(-0.71)					
Distance-1 Density		0.0364				
		(0.48)				
Group Size $ imes$ Distance-1 Density		-0.00769				
-		(-0.41)				
Distance-2			0.0130			
			(0.25)			
Group Size $ imes$ Distance-2 Density			-0.00513			
			(-0.42)			
Distance-3 Density				0.0377		
				(0.71)		
Group Size $ imes$ Distance-3 Density				-0.00763		
				(-0.65)		
Community Density					-0.0362	
					(-0.74)	
Group Size $\times$ Community Density					0.00930	
					(0.86)	
N	526	526	526	526	526	
Session FE	Yes	Yes	Yes	Yes	Yes	

Table F1: Defaults by Group using Colombia Friends and Family Network

t statistics in parentheses

	Proportion of Defaults in Risk Sharing 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 $\times$ Supported Density	0.0293					
	(1.90)					
Distance-1 Density		-0.155**				
·		(-3.25)				
Group Size $ imes$ Distance-1 Density		0.0390*				
		(2.61)				
Distance-2 Density			-0.118*			
			(-2.46)			
Group Size $\times$ Distance-2 Density			0.0323*			
			(2.61)			
Distance-3 Density				-0.114*		
				(-2.40)		
Group Size $ imes$ Distance-3 Density				$0.0302^{*}$		
				(2.42)		
Community Density					-0.120*	
					(-2.39)	
Group Size $\times$ Community Density					$0.0282^{*}$	
•					(2.15)	
N	526	526	526	526	526	
Session FE	Yes	Yes	Yes	Yes	Yes	

Table F2: Defaults by Group using Colombia Close Friends and Family Network

t statistics in parentheses