Social Network Structure and the Scope of Risk Pooling

Daniel S. Putman *

This Version: April 9, 2021
Click Here For Most Current Version

Abstract

The scope of risk pooling refers to the set of individuals one can pool risk with and encapsulates both the size and diversity of the pool. In practice, while risk sharing transfers are mediated by bilateral networks, the network neighborhood may not serve as an ultimate measure of the scope of risk pooling. With laboratory experiment data from Colombia, I explore how social network structure drives the formation of experimental risk pooling groups. Using dyadic regression, I find that direct connections, supported connections, and detected community co-membership explain co-membership in experimental risk pooling groups. In addition, the combination of these measures can detect strong and weak ties when given only an unweighted network. This work has implications for the welfare derived from risk pooling, non-market spillovers, the collection of social network data, and the assumptions commonly used within the risk sharing literature.

Keywords: Risk & Uncertainty, Risk Pooling, Group Formation, Network Formation, Experimental Games, Community Detection

JEL Codes: O12, 017, L14, D85

*Postdoctoral Fellow, Innovations for Poverty Action. dputman@poverty-action.org. This paper has benefited immensely from conversations with Travis Lybbert, Ashish Shenoy, Laura Meinzen-Dick, Arman Rezaee, and Cathy Putman. All errors herein are my own.

1 Introduction

Risk pervades the economic lives of the poor: it determines 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 (Karlan et al., 2014; Elbers et al., 2007). Similarly, vulnerability to uncertainty itself reduces welfare in an *ex ante* sense (Ligon and Schechter, 2011). Despite this, formal financial markets to deal with risk, including insurance markets, are often missing (Mccord et al., 2007). In the absence of formal insurance markets, informal risk pooling built on trust and reciprocity is a common and important method of managing risk (Karlan et al., 2009). Trust and reciprocity are powerful but limited tools to ensure cooperation. As the size and diversity of risk pooling groups grow, it becomes more difficult to rely on trust. However, good risk pooling groups are both large and well diversified. The size of the pool matters because individual shocks vanish in the average as group size grows. Likewise, the diversity in occupation, geography, and other resources matters. When only idiosyncratic risk is pooled, this diversification ensures that risk faced by individuals is idiosyncratic to the group. In this sense, the more diverse the risks that are faced are, the great the proportion of risk individuals can pool. The *scope* of risk pooling refers to the relevant set of individuals one pools risk with. This encapsulates these two major features of risk pooling: the size and diversity of the pool.

Given the importance of risk pooling and its common application as a tool to manage risk in developing countries, understanding how large and diverse risk pooling groups are is a natural question. Despite this, no attempt has been made to catalogue and empirically compare the candidates for the scope of risk pooling presented in the literature. To address this gap, in this paper, I answer the question: what is the scope of risk pooling in networks? My approach in answering this question starts with measuring how relationships are embedded in networks, and how these networks can explain experimental formation of risk pooling groups. Hence, a first question of exploration is who shares risk with who, and what features of network structure determine this decision?

The data combines real world friends and family networks collected via survey methods with an incentivized risk pooling lab experiment Attanasio et al. (2012b). I estimate an econometric model of network formation using dyadic regression to test whether measures of social network structure can explain selection into these experimental risk pooling groups. That is, I explain behavior in the risk pooling experiment using real world network structure. This approach treats the dyad as the unit of observation, thus the outcome of the regression is co-membership in a risk pooling group. This allows me to determine which measures of social network structure are useful to explain who pools risk with whom in the experiment. This social network structure, in turn, can then help inform the scope of risk pooling. For example, a person may be more likely to share risk with someone who is close to them in their social network. This might mean a direct connection, but this relationship may be stronger if their relationship is "supported" by a third friend or family member who observes both in the relationship (Jackson et al., 2012). If this measure explains co-membership, but other measures do not, we should expect a much smaller scope of risk pooling than networks would otherwise indicate. Candidate measures of social network structure include these supported connections, direct connections (Fafchamps and Lund, 2003), distance-2 ("friends of friends") and distance-3 ("friends of friends of friends") connections (Bourlès et al., 2017; de Weerdt and Dercon, 2006), and co-membership in detected communities. I argue this final measure is a good proxy for the features that promote good risk sharing in networks as is detailed in Bloch et al. (2008) and Ambrus et al. (2014). To add context to these estimates, I analyze how defaults in groups are related to network structure.

Using estimates from the dyadic regressions, measures indicating closer social proximity translate more strongly into risk pooling in the experiment. While direct connections consistently explain co-membership in experimental risk pooling groups, distance-2 connections explain co-membership less consistently. On the other hand, I do not find that distance-3 connections help explain co-membership the same experimental risk pooling group. Additionally, supported relationships and detected community co-membership are also help explain co-membership in experimental risk pooling groups. These final two measures capture tightly knit social network structure, albeit at different scales. In particular, support documents the presence of a third friend or family member who is also connected to both members of the dyad. Community detection on the other hand, scales this to capture large, densely connected and clustered groups in social networks. Dyads who are co-members of one of these detected communities will tend to have common friends even if they are not friends themselves.

Estimates of regressing default rates on various measures of network density do not indicate that these measures of network structure are related to default. This however may be due to the inability to punish friends and family who do not live close by. Replication using the close (geographically proximate) friends and family network associates lower default rates with higher network density.

That distance-2 and community co-membership matter for the eventual experimental risk pooling groups imply a scope of risk pooling in this setting that extends beyond one's immediate network neighborhood. Likewise, the failure of distance-3 nodes to explain joining the same risk pooling group suggests risk pools at a much smaller level than the village.[1] This is at first counter-intuitive. Very often improvements in measurement simply make clear to the econometrician what the subjects of the econometric study already understand. For example, when we measure

[1]This falls well short of the diameter of the "giant components" of these networks.

risk sharing networks, we elicit what respondents already know about their networks. However, measurement using community detection often pairs people who are not aware that they lie within each others risk pooling groups. In this way, the subjects of the study may not fully appreciate their own risk pool beyond their network neighborhood.

Potential ties outside of the network neighborhood tend to be (relatively) weak. That is, the probability of co-membership in an experimental risk pooling group falls if we consider community co-membership relative to direct connections. Despite this fact, community co-membership also helps understand relationships within the network neighborhood. In particular, if two agents are supported by a third agents and also are co-members of a community, they have around a 24pp excess probability of matching, relative to an 8pp excess probability of a similarly supported pair who are not co-members of a community.This is expressed as *excess* probability of matching since this analysis controls for municipality level fixed effects, and thus it is in excess of those. In this way, when we looking at the explanatory nature of the supported relationships, we see an added propensity to connect with a subset of agents smaller than even the level of supported relationships. These relationships serve as strongest ties on the extensive margin. In this way, we can build a set of strong and weak ties in social networks in a natural way using only one network. This is valuable to researchers who want to collect different intensities of relationship with a relatively condensed network survey.

Finally, we can use these results as a to provide an audit on the state of economic models of risk pooling. Because risk sharing networks are difficult to measure, the literature is often driven by theory. All authors are left to make an assumption on the scope of risk pooling. At a high level, authors have three choices: risk pooling happens bilaterally, in a network setting, in a group setting, or in the village setting. I review the literature on risk sharing and risk pooling, explicitly documenting assumptions and results about the scope of risk pooling. This meso-level (from the perspective of the village) scope of risk pooling adds credence to approaches that model risk pooling at the subvillage level in groups and/or networks, including work in Genicot and Ray (2003), Bloch et al. (2008), and Ambrus et al. (2014).

To test the robustness of the results, I repeat the above analysis with a close friends and family network, which restricts the friends or family to those dyads who have geographically proximate dwellings. These results mirror those using the main network. Likewise, to isolate these measures as a meaningful mechanism of group formation I include a battery of measures of affinity that might drive the formation of these risk pooling groups. The results are robust to sums and differences in experimental choices made before the risk pooling experiment, winnings, age, gender, education, household structure, and consumption.

These results are relevant for both policy decisions, evaluation, and design. First, some evidence has shown that development interventions (including increased financial access) may have the unintended consequence of eroding informal risk sharing (Banerjee et al., 2018; Heß et al., 2018; Cecchi et al., 2016; Dizon et al., 2019). Despite the value of increased financial inclusion, understanding this trade-off is important to understanding and measuring overall financial health.[2] Moreover, measurement of the quality of informal financial networks (e.g., insurance provided) is often an amorphous and difficult task. An important step in doing so is measuring the scope of risk pooling. To this end, an early set of papers assume the scope of risk pooling occurs at the village level. . However, given these results, this assumption should overstate the welfare gains from risk sharing. More recently, network models have been considered, often viewing risk sharing as a purely bilateral process. It is not clear that risk pooling is simply bilateral, in fact, the empirical results herein suggest that this is a conservative assumption in terms of the value of risk pooling.

Understanding the scope of risk pooling can help us understand the scope of spillovers by non-market mechanisms and thereby is relevant to program evaluation and design. Economists and other social scientists are interested in the scope of spillovers for a number of reasons. The scope of spillovers matters for identification of treatment effects when spillovers confound their estimation. Understanding the scope of risk pooling could help development economists who hope to control for these effects. Additionally, sometimes spillovers are considered as a component of policy design. Both of these issues may be relevant in contexts where it is difficult to deliver cash or in-kind transfers, or simply because of knowledge that these types of spillovers will happen. Considering cash transfers, for example: while some of the spillovers from cash transfers are mediated through market mechanisms, a sizeable portion may come through informal transfers.[3] Risk pooling involves transfers of cash, food, and other commodities in response to various positive and negative shocks. When such a transfer does come as a shock (e.g., during an RCT), these types of spillovers could cloud the estimation of treatment effects. However, the knowledge of the scope of informal spillovers could complement strategies such as those in Leung (2019a) and Leung (2019b), which study the estimates of treatment effects in the presence of network mediated spillovers. Likewise, in the case of program design, knowledge of the scope of risk pooling would inform the scope of pass-on treatment.

[2]For an example of a project with such goals, see Karlan and Brune (2017).

[3]See also Heifer International's Pay-it-Forward mechanism (Janzen et al., 2018).

2 Literature Review

2.1 The Second Best World of Risk Sharing

Evidence of imperfect risk pooling abounds at the village level (Townsend, 1994; Kinnan, 2019; Ligon, 1998; Chiappori et al., 2014). On the other hand, both the theory and evidence of informal risk pooling points to it's scope being well within that of the village since in the absence of clearinghouses, bilateral transfers are the mechanism by which risk sharing tends to takes place (de Weerdt and Dercon, 2006; Fafchamps and Lund, 2003; Collins et al., 2010). Still it is unclear what the full scope of risk pooling within the village is in practice. An aggregate measure of risk sharing does not differentiate between imperfect risk sharing on the intensive (e.g., proportion of risk shared by each individual) as opposed to the extensive margin (e.g., the size of risk pooling groups). Likewise, higher order (e.g., distance-2) connections are often ignored.

Efficient risk sharing arises as a natural benchmark to compare to any observed outcomes. Diamond (1967) models how contingent commodity markets (which resemble informal risk sharing without any market imperfections) can achieve optimal outcomes.[4] In the context of pooling idiosyncratic risk, optimal outcomes suggest all idiosyncratic risk is pooled at the village level. However, in practice, while risk sharing tends to occur, it is rarely efficient. Townsend (1994) develops an econometric model to test for the presence of efficient risk sharing at the village level. This model suggests that if risk is being shared efficiently, idiosyncratic shocks should not influence individual consumption if aggregate consumption has been controlled for.[5] The paper investigates this hypothesis using data from southern India and shows evidence that while idiosyncratic risk is shared at the village level, risk sharing is incomplete. To explain the failure of village economies to achieve complete risk sharing, many strands of literature have emerged. These explanations include (but are not limited to) hidden income and assets (Baland et al., 2011; Kinnan, 2019; Cabrales et al., 2003), moral hazard over risk and effort (Boucher and Delpierre, 2014; Delpierre et al., 2016; Kinnan, 2019), transaction costs (Jack and Suri, 2014), and limited commitment (Coate and Ravallion, 1993; Ligon et al., 2002; Bloch et al., 2008; Kinnan, 2019; Ambrus et al., 2014). All of these serve to place constraints on bilateral risk sharing, or risk pooling at the village level, through information asymmetries.

[4]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.

[5]Note that this doesn't mean that income or wealth will have an egalitarian distribution. In fact, in Fafchamps (2008) the opposite is suggested: idiosyncratic risk sharing should roughly preserve the ordering of incomes.

2.2 Matching Measures to the Literature

We can organize the risk pooling literature by assumptions made about the scope of risk pooling. In this section of the paper, I match various definitions of scope to the assumptions made in the literature. The literature focuses on important work in the field of risk sharing and includes definitions used in theoretical and in empirical work. The definitions will in turn drive the measures I use in the empirical work presented in the sections below.

One way to define the risk pooling groups to use the network neighborhood (or the set of agents one is directly connected to) is the scope at which agents pool (Fafchamps, 1999; Fafchamps and Gubert, 2007; Jack and Suri, 2014; Blumenstock et al., 2016). Second, where informal or quasi-formal groups are identifiable (e.g., burial groups), we can define the scope of risk pooling at this level (Fitzsimons et al., 2018; Dercon et al., 2006). Finally, we can model risk sharing at the network level, allowing the scope of risk pooling to depend on network structure. For example, friends of friends might be part of an agent's risk pool if transfers tend to flow through networks (Belhaj and Deroïan, 2012). On the other hand, due to enforceability concerns (for example), members of the network neighborhood might be excluded if they don't have common friends (Jackson et al., 2012; de Weerdt, 2002). In the following subsections, I examine these different definitions in depth.

2.2.1 Bilateral Risk Pooling and Network Neighbors

Many studies take the network neighborhood, or the set of individuals directly connected to an agent as the relevant unit of risk pooling (de Weerdt and Dercon, 2006; Fafchamps and Lund, 2003; Collins et al., 2010; Fafchamps, 1999; Fafchamps and Gubert, 2007; Jack and Suri, 2014; Blumenstock et al., 2016; Jack and Suri, 2014). Hence, this is our first measure of the scope of risk pooling. We can define this pool as the set of individuals who are directly connected to agent i in the network:

$$N_i(g) = \{j|ij \in g\} \tag{1}$$

This scope of risk pooling assumes a number of things about the data at hand, which will be relaxed as we move into further measures. First, it assumes that flows on networks do not matter. That is, if i transfers to j, this transfer will not continue to flow through the network to other neighbors of j. Second, it imposes that the network is static, i.e., that no links will be systematically removed or added thereafter. Third, we are assuming away that friends common to both i and j do not matter. This would exclude cases like those studied in the literature of limited commitment, where this common friend serves as additional incentive not to renege.

2.2.2 Distance-s Connections

We first consider the case where people in the network who are more distant matter for risk pooling. To do this, we build a unit of risk pooling that includes all agents up to distance of s steps between nodes away from agent i. I define a shortest paths distance-s neighborhood, similar to the neighborhood definition above. Where the network neighborhood involves all agents who can be reached in one step, the distance s neighborhood includes those who can be reached in a minimum of s steps. This is defined

$$N_i^s = \{j | \min distance(i, j) = s.\} \tag{2}$$

A visualization of the distance-s sets of nodes is presented in Figure 1. The rationale for higher distance connections might be important because of unobserved first order connections, network dynamics (e.g., introduction by friends), or flows on networks.

Unobserved first order connections and network dynamics are closely related. Starting with network dynamics, there are cases where the assumption that only direct connections matter may be be perfectly sensible. For example, if the data collected is forward looking and we are interested in risk sharing shortly thereafter ("who would you ask..." questions) and agents don't have an incentive to misreport, this static assumption may make perfect sense. However, in *ex post* data sets ("who have you transferred with..."), we might expect that not all possible risk sharing partners have been exercised previously (asking for a favor is costly). Likewise, we also might expect agents to meet new friends, often through introduction through existing friends. Hence, connections that are relevant in the future may not be in the defined set. This possibility is particularly relevant to the risk pooling experiment in our case study. Additionally, many networks in the real world tend to be clustered. That is, it is likely that when i connects to j and j connects to k, i also connects to k. In highly clustered networks, distance-2 connections may be a good place to look for omitted connections.

While not relevant to this empirical setting, network flows are an interesting possibility.[6] For example, Bourlès et al. (2017) builds a theoretical model of altruism in networks (related, though distinct from purely self-interested risk sharing) and find that intermediaries are important to flows of transfers through networks. In their model, cases where $ij, jk \in g$ but $ik \notin g$, j may serve as an intermediary. Supposed i is altruistic toward j and j is altruistic toward k, the intermediary j requests a transfer from i in order to make a transfer to k, who is in need.

[6]Note that this is because risk pooling groups must be joined explicitly in the risk pooling experiment. These flows of transfers are "shut off" in our observation of risk pooling behavior. However, since additional transfers may take place after the experiment ends, we can't rule the importance of flows out entirely. Specifically, the ability to insure though network flows may trade off with more costly connections. If we were to rule these out, knowledge that flows will not take place might induce a second degree connection, for example.

Another possibility when network flows are introduced happens when many transfers are taken. If one pools risk with one's immediate neighborhood, but is given unlimited costless transfers to deal with risk, risk pooling groups could extend far beyond that network neighborhood (this is proven in Bramoullé and Kranton (2007)). Despite the possibility of network flows, de Weerdt and Dercon (2006) find that direct connections matter for illness related risk sharing, but distance-2 connections do not.

2.2.3 Support: Common Friends

Second, we can examine the case where a smaller subset of the network neighborhood matters more for risk sharing than the average connections. In particular, where a common friendship exists, we can define a supported neighborhood. We can define the supported neighborhood as follows:

$$SN_i(g) = \{j | ij \in g,\ \exists k, jk, ik \in g\} \tag{3}$$

We can see a visualization of this set is presented in Figure 2. As described above, this definition involves a restriction on the neighborhood measure we've already defined above. Here, while j must already be within agent i's neighborhood, we only count them as within the scope of risk pooling if there is some third agent k observing the interaction. This agent acts to support relationship between i and j. In particular, this approach is used in Jackson et al. (2012) to model favor exchange in Indian villages; they find that stable networks are those where links are supported by an observing node. In risk sharing, more specifically, support ties itself to problems of asymmetric information. For example, limited commitment in risk sharing networks occurs when contracts on networks cannot be enforced. Therefore, relationships need to be self-enforcing. Punishment for reneging on an obligation to transfer money to one's worse off neighbor often (in theory) means being ostracized by those observing (Coate and Ravallion, 1993; Ligon, 1998). We expect to see stronger risk sharing relationships when support is present.[7]

2.2.4 Subvillage Risk Pooling Groups: Found and Detected

Third, we can consider the case where larger risk sharing groups exist in the village and matter for risk sharing. The members of these groups may not be friends, but particularly when groups are formed with the purpose of insurance matter for risk pooling. Similarly, we may also be able to recover latent risk sharing groups from network data when natural grouping of agents often

[7] Generalizing from this idea, we might expect to see stronger relationships the more supporting nodes exist, though our measure of support does not account for this.

exist (Pons and Latapy, 2004). We define an agent's *community* as follows:

$$CN_i(g) = \{j | j \neq i,\ i, j \in C_i\} \tag{4}$$

where C_i is a given group determined either by actual group membership (i.e., a funeral insurance group) or by processing the network somehow. For the first possibility, many examples exist of "found" informal risk sharing groups. For example, funeral societies in Ethiopia and Tanzania (Dercon et al., 2006) and kinship groups in Malawi (Fitzsimons et al., 2018) are both examples of group level risk sharing derived from institutional context. However, contexts where we can determine these groups ex ante tend to be the exception. In other contexts, however, it is reasonable that informal groups might still exist. For example, Murgai et al. (2000) uses an intuitive coding of clusters along irrigation canals in Pakistan and shows that insurance related water exchanges in this context happen among households within these tightly knit clusters. Additionally, following with a common pattern of results, the paper looks at the optimal cluster and finds an extensive/intensive margin trade-off.

In the theoretical literature, it's valuable to review the literature on the stability of risk sharing groups and networks, and the resulting characteristics of those groups. Genicot and Ray (2003) explores the formation of risk pooling groups with limited commitment using a theoretical model. Groups which are stable (in the sense that they are self-enforcing) are bounded in size. Bloch et al. (2008) can be thought of as an extension of this work, examining the stability of risk sharing networks.[8] In this case, networks must act as conduits transfers and also information. While the details of the paper are technical, the authors find certain network structures that are easy to pass information in, which in turn makes punishment more effective when someone reneges. These network structures turn out to be some with with low density such as trees and lines and others with high density such as the complete graph or a "bridge" graph. This final example of the "bridge" graph, a set of two small cliques[9] connected by one bridging link, is very relevant here because it provides rationale for network structure that closely accords with community structure. Finally, Ambrus et al. (2014) builds a theoretical model of the effect of network structure on *ex post* consumption risk sharing. They find that commonly observed network structures do not imply complete (optimal) risk sharing. Moreover the case of incomplete (second best) risk sharing, after shocks have been realized, they hypothesize the emergence of risk sharing "islands" 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.

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

[9]A clique is a complete subgraph occurring in a network graph.

The results of these three papers, combined and individually motivate a search for searching for groups within networks. There are a number of ad hoc ways we might get at the network structure suggested in this literature. First, as Bramoullé and Kranton (2007) suggests, we could look at components of risk sharing networks. In this case, we would include any individual where a path exists as part of the network in our group. However, this seems to ignore the emphasis on network structure found in both Bloch et al. (2008) and Ambrus et al. (2014). Likewise one could search for cliques (completely connected subgraphs) within networks. This may be useful, but will be highly correlated with measures of supported communities. Furthermore, this definition is highly inflexible to the structure of network data.[10] Finally, my third and preferred option is to use clustering algorithms to detect likely communities. These communities lack the clean definition of the clique or component but have the advantage of being able to tame messy data into a consistent unit. For a visualization of community detection, see Figure 3.

3 Data and Context

3.1 The Experiment

The data come from a laboratory experiment in Colombia and were obtained as replication files from (Attanasio et al., 2012b).[11] In addition to experimental results, the data features real world social networks and a rich set of demographic variables. The empirical work follows a basic structure in using this data: how well can we explain experimental decisions with real world network data? That is, experimental measures serve as outcomes while survey measures serve as explanatory variables. In this section, I briefly explain the risk pooling experiment, sampling, and recruitment, as well as the real world social networks survey measures.

The risk pooling experiment was conducted in 70 Colombian municipalities in 2006. The experiment collected information both about risk preferences and risk pooling groups in two rounds of play. The first round of play consisted of a gamble choice game. This was followed by period where individuals were allowed to talk and form risk pooling groups to share their winnings from a second gamble choice game. Finally, individuals played a second gamble choice game and winnings were distributed according to the risk pooling groups formed.

[10] For example, consider an "almost-clique" which is missing just one connection. Is it more natural to think of this as two cliques, or should the two unconnected agents who have many friends in common provide insurance for each other?

[11] It is worthwhile to note that prior to starting this research project, I was able to replicate the results of Attanasio et al. (2012a) in a push-button replication.

3.1.1 The Gamble Choice Game

The first round of the risk pooling experiment consisted of a version of the Binswanger (1980) gamble choice game. In this round the the experimental subjects chose one gamble from a list of six presented to them. As can be see in Table 2, 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 the subject chose their gamble, they played the gamble of their choice and received a voucher for their payout.

3.1.2 The Risk Pooling Game

Round two of the experiment consisted of another gamble choice game, but this time with the opportunity to pool the risks of this game. This time, before meeting with the experimenters, the subjects were allowed to form risk pooling groups in which winnings would be pooled and shared equally. Subjects were given around an hour to an hour and a half (during lunch) to form their groups. These groups were declared before the second set of meetings took place. During the meetings, subjects were given the chance to privately withdraw from their groups after seeing the outcome of their gamble. Subjects were informed of this fact before they started to form they groups. In this case, when they withdraw, they forfeit their share of the group earnings and not need to share any of their earnings either. The remaining group members would pool their gambles and share these equally. Thus, one's earnings depend on the size and composition of the group after any withdrawal.

Letting $\ell \in \{1, \ldots, 6\}$ be an individuals type, then 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

$$E(y) = \sum_{\ell=1}^{6} q_\ell \times E(y_\ell) \tag{5}$$

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

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

and N_G is group size. Considering withdrawal once again, it is only rational for and individual not to withdraw from the risk pooling group unless their revealed income exceeds the expected income. In reality, both the expected income above and the variance of the mean should shrink. Given subject expectations about withdrawals, we could re-write each of these expressions based on expected composition of groups after withdrawal (i.e., the output of a game theoretic model).

3.1.3 Sample and Recruitment

These 70 municipalities were drawn from a sample of 122 that were surveyed to evaluate Colombia's national cash transfer program Familias en Acción. In the sampled municipalities, about 60 households were invited to an experimental workshop in their municipality. These families reside in the poorest sixth of the national population and the attending members were largely female (as they are the recipient of the cash transfer and hence received the invitation). In total, 2512 individuals took part in the experiment.

3.1.4 Summary of Experiment Outcomes

86.9% of subjects chose to join a risk pooling group. These groups tended to be small, with an average of 4.6 members. 6.4% of subjects defected from their group after winning the second round gamble.

3.2 Network Data

3.2.1 Data Collection

Networks are collected on the day of the experiment. This takes place by asking each participant in the experiment if they know others and to clarify the nature of the relationship (family or friend). Since the experimental subjects are sampled from a larger population of interest, collecting them on the day of the experiment means that the networks are sampled. Note that while Chandrasekhar and Lewis (2016) warns of the pitfalls of doing regression with network statistics using sampled networks, node sampling is a standard assumption in dyadic regression.[12] With that said, it is less clear how network sampling impacts measures like supported connections or second order connections. This is an important area of analysis to fully understand the results.

3.2.2 Social Network Summary

Details of the social networks characteristics are presented in Table 3. Experimental groups vary in size by municipality, ranging from 9 to 87 experimental subjects. On average, around 34 subjects were invited to attend. The friends and family networks tend to be sparse, with an average density of 5.6%. This indicates that of all potential connections (within the municipality), only about 1 in 20 exist. The networks are moderately clustered with a clustering coefficient of 34.6%. This means that an agent in the network knows one third of the people in their neighborhood's

[12]See for example, Graham (2020).

connections.[13] Closeness in the data is 0.55, suggesting we can think of the average distance between dyads to be around 1.8 steps in these networks.[14] Using the Walktrap community detection algorithm, we see detected communities of average size 3.93.

4 Empirical Strategy

4.1 Dyadic Regression Specifications

To test the explanatory power of various measures of the scope of risk pooling, I use dyadic regression, an econometric model of network formation. In these regressions, each pair of agents (i.e., a dyad) is treated as an observation. Therefore, I translate the measures used in 2 into dyadic measures of network proximity. Based on the structure of the data, I only include dyads that were in the same municipality, since the meetings for the field experiment took place at the municipality level.

4.1.1 Main Specification

We start with a simple specification that seeks to explain co-membership in risk pooling groups using friendship or family ties in social networks. All dyadic regression specifications share a common outcome: co-membership in the experimental risk pooling group. That is, in the incentivized experiment, do agent i and agent j decide to join the same risk pooling group? On the other hand, the right hand side regressors are all constructed from real world social networks. Hence, we seek to explain dyad level behavior in the experiment based on subjects' real world social networks. From the results of Attanasio et al. (2012a), we already know individuals tend to join the same experimental risk pooling group if they are (geographically proximate) friends or family. However, other measures of social proximity are not tested. Hence, I seek to test other measures of risk sharing in comparison to the importance of friends and family.

If some additional measure is to add value above direct connections, it should be able to explain variation in co-membership in experimental risk sharing groups above and beyond these other measures. The first set of estimates focuses on three kinds of dyads: friends and family, supported relationships, and co-membership in a detected community. While detailed descriptions of these dyads are presented in 4.2, I summarize them here. First, a dyad are friends or family if

[13]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?

[14]More precisely, closeness is computed by taking the average of inverse of shortest path distance for nodes in the network. When the distance is ∞, we take closeness to be 0. 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.

both recognize friendship or family ties. Second, this relationship is supported if there is a third agent who is in the network of a connected dyad. Finally, a dyad are co-members in a detected community if both belong to the same detected community. The main specification is

$$Group_{ij} = \alpha_v + \beta_0 S_{ij} + \beta_1 A_{ij} + \gamma C_{ij} + \varepsilon_{ijv} \tag{7}$$

where α_v is a municipality fixed effect, $Group_{ij} = 1$ indicates i and j joined the same experimental risk pooling group, $S_{ij} = 1$ indicates a supported relationship, $A_{ij} = 1$ indicates a friend or family tie is present, and $C_{ij} = 1$ indicates 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 join a risk pooling group. On the other hand, $\gamma > 0$ indicates that perhaps not everyone in the risk pooling group falls within the network neighborhood.

4.1.2 Longer Walks: Increasing the Radius of Risk Pooling

While detected communities may be one way we see increased scope of risk pooling, it may be that anyone within a specific radius is important for risk sharing. To test this, I include dummies for those dyads who are 2 and 3 steps from each other. To test this, I include these "longer walks" on their own as as with measures of support and community. This specification can be written

$$Group_{ij} = \alpha_v + \beta_0 S_{ij} + \sum_{s=1}^{3} \beta_s A_{ij}^s + \gamma C_{ij} + \varepsilon_{ijv} \tag{8}$$

where $A_{ij}^s = 1$ indicates there is a shortest path of length s between i and j. In addition to the tests detailed above, here I test $\beta_s > 0$ $s = 2, 3$. Tests of β_s indicate (similar to tests of γ) that perhaps not everyone in one's risk pooling group falls within the network neighborhood. If rejected, these tests will indicate that those further flung members in one's social network are good candidates for pooling risk. However, since community co-membership and distance are closely related, the correlation when accounting for this measure is likely more meaningful. In terms of the magnitude of these effects, qualitatively, I would expect that closer dyads are weakly more likely to match, i.e., $\beta_1 \geq \beta_2 \geq \beta_3 > 0$.

4.1.3 Fully Interacted Specification

Finally, there is a great deal of heterogeneity in the dyads of agents who are co-members in communities, including their distance apart and whether or not their relationship is supported by a third agent. 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-membership in risk pooling group 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:

$$Group_{ij} = \alpha_v + \beta_0 S_{ij} + \sum_{s=1}^{3} \beta_s A_{ij}^s + \gamma C_{ij} + \delta_0 S_{ij} C_{ij} + \sum_{s=1}^{3} \delta_s A_{ij}^s C_{ij} + \varepsilon_{ijv}. \tag{9}$$

In addition to previous hypotheses, I expect dyads within communities (at a given distance) are more like to match than those dyads between communities. That is, I test $\delta_s > 0$ $s = 0, \ldots, 3$. In addition, excess probability of co-membership can be estimated for each of nine dyad types (relative to dyads who are not connected, supported, or co-community members). For computation for each of these dyad specific means, see Table 4.

4.2 Variable Construction

4.2.1 Co-Membership in Experimental Risk Pooling Groups

The outcome of interest in the dyadic regression is whether or not a dyad of individuals joined the same experimental risk pooling group. Being in a risk pooling group with the other member of the dyad is referred to as co-membership in the risk pooling group. Note that groups are non-overlapping: each agent can only be in one group. For $i \in Group_i$ and $j \in Group_j$, we define $Group_{ij} = \mathbf{1}(Group_i = Group_j)$.[15]

4.2.2 Friends and Family Network

The explanatory variables of interest are constructed from network survey data. In contrast to the experimental risk pooling groups, this is a network of *real world* friendships and family ties. We start by forming an undirected and unweighted friends and family graph g. We say $ij \in g$ if either i recognizes j as a friend or family member or j recognizes i . For a graph g, I define the adjacency matrix $A_{ij} = \mathbf{1}(ij \in g)$. For second order connections, we look for friends of friends (or family of friends, friends of family, etc.). In terms of graph theory, we define these second order connections as all dyads that have a minimum distance of 2 between them $A_{ij}^2 = \mathbf{1}(\min distance(i,j) = 2)$ where distance is defined as the number of steps one counts when

[15] I use an indicator function defined

$$\mathbf{1}(\text{condition}) = \begin{cases} 1 & \text{, if condition is true} \\ 0 & \text{, if condition is false} \end{cases}. \tag{10}$$

traveling over edges between nodes.[16] Third order connections are defined as any dyad with a shortest path of 3: $A^3_{ij} = \mathbf{1}(\min distance(i, j) = 3)$. Finally, for supported relationships, we take any dyad who has a third friend in common. Using the graph theory representation, supported relationship are defined as $S_{ij} = \mathbf{1}(ij \in g$ and $\exists k$ such that $ij, jk \in g)$. For the close friends and family network, these definitions still apply, we simply reconstruct g. In particular, we restrict $ij \in g$ to only occur if both i and j both recognize friendship or family ties and i and j also are geographically proximate to each other.

4.2.3 Co-Membership in 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, g. Specific approaches for this assignment are discussed in detail below in Section 4.4, but I can define community co-membership with just an understanding of the eventual assignment. In particular Each agent is assigned to exactly one community (this includes communities that feature only one member, in some cases). We say $i \in C_i$ and $j \in C_j$. Then we define community co-membership as $C_{ij} = \mathbf{1}(C_i = C_j)$.

4.3 Group Specifications: Network Structure and Defaults

4.3.1 Econometric Specification

Are networks tied to the rate of defaults within experimental risk pooling groups? To provide context, I build on the analysis from Appendix Table A1 of Attanasio et al. (2012a). In particular, at the group level, I estimate the following specification:

$$Pr(Default|G, v) = \alpha_v + \beta N_G + \gamma Density(G, .) + \delta N_G \times Density(G, .) + \theta \bar{X}_G + \varepsilon_{Gv} \quad (11)$$

where N_G is the size of group G, $Density(G, .)$ is one of a set of network densities, $\bar{X}_G$ is a set of controls consisting of group means, and α_v are municipality level 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.

[16] The minimum distance is the number of steps one would have to take through the graph to travel from i to j. We could also write this more laboriously as $A^2_{ij} = \mathbf{1}(ij \notin g$ and $\exists k$ such that $ik, kj \in g)$. Indeed, second and third order connections are computed similarly in the data.

4.3.2 Variable Construction

I use five definitions of network density: supported density, distance-1 density, distance-2 density, distance-3 density, and co-community density. Density is computed by taking in the number of dyads within the group with the given characteristic (e.g., "are connected") over total number of dyads within the group. More formally, I compute supported density[17] as

$$Density(G, S) = \frac{\sum_{i,j \in G} S_{ij}}{2N_G(N_G - 1)}, \tag{12}$$

distance-1 density as

$$Density(G, A) = \frac{\sum_{i,j \in G} A_{ij}}{2N_G(N_G - 1)}, \tag{13}$$

and community density as

$$Density(G, C) = \frac{\sum_{i,j \in G} C_{ij}}{2N_G(N_G - 1)}. \tag{14}$$

Distance-s density generalizes network density, includes all dyads of minimum distance less than s.[18] More formally I compute Distance-s density,

$$Density(G, A^s) = \frac{\sum_{i,j \in G} \sum_{t=1}^{s} A_{ij}^t}{2N_G(N_G - 1)}. \tag{15}$$

4.4 Community Detection

I propose two community detection algorithms in the following section and argue for their relevance to risk sharing: edge betweenness and walktrap.[19] The first of these two approaches is the edge betweenness algorithm from Girvan and Newman (2004) and the second is the Walktrap algorithm from Pons and Latapy (2004). In practice the two do not differ meaningfully, so I use the Walktrap algorithm.

4.4.1 Edge Betweenness

Betweenness centrality is often used as a measure of the potential for brokerage and information bottlenecks. Working from the intuition of information bottlenecks presented above, I propose

[17] Why $\frac{1}{2}$? Because summing over all entries of the relevant adjacency matrix double counts the number of connections.

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

[19] While it is perhaps intuitive to approach finding these communities using a direct approach, this is difficult to come by. For example, the approach described in appendix suffers in it's ability to differentiate unions of small risk pooling groups from larger risk pooling groups.

edge betweenness community detection. The algorithm calculates the betweenness centrality of all edges in the network and deletes those edges with highest centrality.[20] This would create a specific rationale for these communities as risk sharing units, then. The types of information bottlenecks that exacerbate barriers to risk pooling are the same places the algorithm cuts ties.

At the start of the process, all nodes in a particular component are assigned to the same community. Every time a component breaks into two with the deletion of an edge, the community's membership is reassigned for the nodes in the broken component. After initially computing all edge betweenness centralities, the algorithm works in a repeated two step process. First, delete the edge with the highest betweenness centrality. Second recompute the betweenness centrality of remaining edges. This two step process continues until all edges are deleted and thus all nodes reside in their own community. To choose a final community assignment, the algorithm cuts the resulting tree of cuts using modularity as a tuning statistic (Girvan and Newman, 2004).[21] See Appendix 6.2.2 for more details on Modularity's computation.

4.4.2 Walktrap

Where edge betweenness relies on intuition about bottlenecks of information flows on networks, walktrap mimics the flow of transfers on networks. Consider a potential risk sharing process. A large gift is given to a randomly chosen household in a risk sharing network. The household gives a gift to a (network) neighboring household who is less well off than they are, sharing their positive shock equally. Having received this gift, this household is also is obligated to share with their worse off neighboring household, provided they are not worse off than all of their neighbors. This process of risk sharing continues. We end up with progressively smaller transfers "walking" randomly through the network. Households who are close to the household who received the prize will have a high probability of receiving a transfer, while those far will have a low probability. Likewise, even if one does not receive a transfer in the first step, if a household is connected to many of the same households as the one receiving the prize, they get additional chances for a gift. We would expect that within a tightly knit portion of the network, most often the gains from the positive shock will not make it far, instead getting "trapped" in the local network.

This intuition mirrors the Walktrap algorithm. A random walker starts at a random node i and moves to an adjacent node with probability $1/d(i)$ where $d(i)$ is the degree of i. This is repeated for s times and the landing node k is recorded. Then nodes are termed similar if controlling for

[20] To compute this measure, the algorithm computes all shortest paths between nodes on a network. Then, count of the number of shortest paths passing through each edge of interest (in the case multiple paths tie for shortest path for two nodes, a partial count is awarded across the edges in these paths).

[21] Modularity is the sum of connections above expectation occurring between individuals within a community. High modularity indicates that communities are tightly knit, so we choose the community assignment with maximum modularity.

landing degree, they tend to land on the same nodes. Each node starts as its own community. Using this measure of similarity, we use a two step process. We merge the most similar adjacent nodes, and then recompute the similarities. This process continues until all nodes are merged into one community. Then, as in the edge betweenness case, we are left with a tree of merges. We arrive at an assignment by cutting this tree at the highest modularity (Pons and Latapy, 2004).

4.5 Robustness

4.5.1 Interpreting the Estimated Coefficients

While understanding the measurement of social networks is of interest, the microeconomics of networks is a field in which the "credibility revolution" often meets practical limitations. 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, in lieu of arguing that network structure causes co-membership in risk pooling groups, I would rather inform the reader of what assumptions would one have to believe to interpret the estimates above as causal.

In particular, in the absence of experimental variation in the network, one might have to rely on a selection-on-observables approach. However, we have at least one advantage from the data at hand: the risk pooling experiment is run after real world networks have been realized. Thus, the interpretation of the estimated results should not suffer from the possibility of reverse causality. To this end, I account for a battery of factors that fall into three broad categories of omitted variable bias: common shocks, popularity, and homophily. However, even after accounting for the factors below, it is important to note that we cannot account for unobservable differences in dyads and/or experimental subjects. It could be that the social network structure and risk pooling membership are the result of these unobserved differences. Thus, one would need to believe that I have accounted for the universe of possible factors.

In any case, as a robustness check, I control for the factors below. First, common shocks may matter, for example, due to experimental conditions. These common shocks might include the execution of experimental protocols during the experiments, which might have varied on the margin. To this end, I include municipality level fixed effects in all regressions to control for municipality invariant features of group formation. Second, we might expect to see that certain individuals are 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 because of their wealth. This effect would manifest itself in both social network structure and choices made in forming experimental risk pooling groups. To control for this type of bias, I include the sum of (log) income, education, risk preferences, and age to control for factors that might drive popularity. Third, we also consider other characteristics that might serve as measures of social distance. Agents who are closer in social, economic, and geographic space tend to be more likely to connect in social networks (Miller McPherson et al., 2001). Hence, I control for the differences in gender, (log) income, education, whether the respondents live in an urban area, risk preferences, and age.

4.5.2 Estimation and Standard Errors

I estimate the above specifications as linear probability models with municipality level fixed effects. To correct for non-independence of standard errors, as in Attanasio et al. (2012a), I cluster at the municipality level. While one accepted approach is to use dyadic robust standard errors (Fafchamps and Gubert, 2007; Cameron and Miller, 2014; Tabord-Meehan, 2019), clustering a natural unit tends to be more conservative and hence is preferred here.

5 Results

5.1 Experimental Risk Pooling and Network Structure

5.1.1 Main Results: Support, Neighborhood, and Community

How well do these measures of the network explain co-membership in experimental risk pooling groups? Across all specifications in table 5, supported friends or family, friends or family, and community co-membership enter positively and significantly (we always reject $\beta_0 = 0$, $\beta_2 = 0$, and $\gamma = 0$ at the 99.9% confidence level).[22] However, the magnitudes of the estimates vary by specification. In particular, the three measures are strongly correlated, and may be picking up some overlapping information about network structure. Hence, I prefer to focus on specification (7), which includes all three. Here, being in the same community is associated with a 6.2 pp increase in joining the same risk pooling group, being in the network neighborhood is associated with a 7.6 pp increase in joining the same risk pooling group, and being in a supported relationship is associated with a 8.3 pp increase in the probability of joining the same risk pooling group.

5.1.2 Longer Walks: Distance-s Connections

While our basic measures of risk pooling seem to do well on their own and in concert, distance-s connections don't entirely fit with this trend. In Table 6, specification (1), distance-2 connections enter significantly (at the 99.9% confidence level). However, the size and significance falls considerably as community dummies enter the regression. The size of the association falls by roughly

[22] This table includes all combinations of the three variables.

half and enters either insignificantly or at significant at the 95% Confidence level. For their part, distance-3 connections enter insignificantly across all specifications.

5.1.3 Interactions of Measures

Specifications (4)-(7) of Table 6 focus on interactions of these various measures with community membership. The general pattern of the results are as follows. First, distance-1 and distance-2 ties tend to enter significantly. First, because distance-1 and supported distance-1 ties are so strongly correlated, this limits the number of significant interactions. My interpretation is that the same set of dyads is reflected in various results in specifications (4)-(7), but where this appears in the regression differs by which measures are used. To cut through the complexity I focus on tests and interpretation of the full specification (7). In this specification, only three terms enter significantly: friends and family (99.9% confidence level), distance-2 (95% confidence level) connections and the interaction of community membership and supported connections (99% confidence level). This yields a striking result: while communities are still important in explaining the formation of risk pooling groups, these work to restrict the set of possible members to the group instead of expand this set of matches.

5.1.4 Strong and Weak Ties

Using a flexible fully interacted regression specification to inspect the excess probability of co-membership in an experimental risk pooling gives us a picture of strong and weak ties in networks. In Figure 6 (based on the formulas in Table 4) we see that not only does the probability of co-membership in a risk pooling group tend to increase the closer one gets in terms of network distance, support, and community membership, but that community membership actually amplifies these other factors. First, as noted above, the shorter the distance is to someone, the higher the probability of co-membership regardless of community status. As seen in the regression tables above, Distance-3 ties tell us little the probability of co-membership, whereas shorter distances are informative. Second, we see that ties within detected communities are (weakly) stronger in explaining risk pooling group formation at every distance.[23] Finally, it becomes obvious that support only seems to matter for within community ties. That is, supported ties across communities are no better than unsupported distance-1 in terms of co-membership in experimental risk pooling groups. In contrast, we see that supported connections within communities (the strongest connections measured).

[23]Weakly stronger in in the sense that we can't always reject the null hypothesis that the means are equivalent.

5.1.5 Network Structure and Defaults

Network density doesn't seem to correlate with group level default rate (at least in the standard friends and family network). The results are both statistically insignificant at any conventional levels of statistical significance and also are tend to be economically small in magnitude and facing in opposite direction of expectation. For example, a 1 pp increase in distance-1 density at the group level corresponds to a 0.036 pp *increase* in the default rate. 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): default fall in network density, but this effect is attenuated as groups grow larger. In terms of success explaining the reduction in defaults, no particular statistic does much better than another, and all are strongly correlated.[24]

5.2 Robustness of Results

5.2.1 Close Friends and Family

One way to check the robustness of the results is to use a different measure of the friends and family network. A reasonable alternative choice is to use only the network of close friends and family. That is, friends and family who are also geographically proximate. Doing this, the main results replicate well. Looking at Table 7, coefficient estimates are a bit larger and noisier, owing to the sparser nature of the close friends and family network.

5.2.2 Kitchen Sink Regressions

Results from the "kitchen-sink" regressions broadly accord with their counterparts. These results can be seen in Tables 9 and 10. For main results, patterns of significance (and rough magnitudes) replicate exactly from Table 5. Comparing Table 10 and Table 6 to examine longer walks and interaction effects, there is not a clear pattern of changes in coefficients, besides adding to the precision of the estimates. For example, while magnitudes change a bit between the two tables in specification (7), Figure 6 can be closely replicated.

[24]It is important to note that is not clear that we would expect to see reductions in defaults to be correlated with density. In particular, one could imagine an theoretical model where groups grow only to a size where very few group members default. This size of course, would be conditional on the underlying network structure.

6 Conclusion

6.1 Discussion

6.1.1 Summary

Using dyadic regression, I explore the explanatory power of measures of network structure in explaining experimental risk-pooling outcomes. In doing this, I specify the scope of risk pooling as conditional on network structure. This allows me to correlate likely measures of risk sharing networks and groups with a "ground-truth" measure of risk pooling. Of these dyadic measures tested, three tend to be particularly useful in understanding the scope of risk pooling: direct connections, supported connections, and co-membership in communities. The third of these measures, depends on community detection, a method that is new to the study of risk pooling. In addition, distance-2 connections also sometimes explain co-membership in experimental risk pooling groups, though these estimates are not stable. However, between community co-membership and distance-2 connections, we see that the scope of risk pooling tends to extend beyond ones direct connections. On the other hand, distance-3 connections consistently fail to explain co-membership in risk pooling groups. Thus

6.1.2 Meso-Level Risk Pooling

These results point toward risk pooling that takes place at a meso-level between the village (or municipality) level and bilateral level. This understanding might guide how we think about the welfare derived from informal risk pooling. For example, we should be wary of any welfare calculations done under the assumption that all members of a village or municipality share risk. On the other hand, models that assume bilateral risk sharing only may be conservative in this regard. When considering the literature of risk pooling, theoretical models that allow for this kind of meso-level risk pooling also become more intriguing. These include, for example Genicot and Ray (2003), Bloch et al. (2008), and Ambrus et al. (2014). Moreover, these results have a special interpretation in relation to Ambrus et al. (2014). In particular, detected risk pooling communities are very related to risk sharing *islands* presented therein. However, they differ in a few important ways. Risk pooling communities map the *ex ante* structure of risk sharing networks while risk sharing islands map *ex post* consumption smoothing conditional on existing networks and realized shocks. This suggests that risk sharing islands arise *ex post* where risk pooling communities exist *ex ante*. The empirical results are consistent with this story. Substituting the experimental risk pooling groups for islands, we see that co-community members tend to join the same risk pooling islands.

6.1.3 Practical Contributions: Strength of Ties and Spillovers

Despite establishing a meso-level of risk sharing, not all network structure is equal. It is still the case that more proximate dyads (in terms of network structure) are more likely to join the same experimental risk pooling group. First, we see that there is a set of individuals smaller than the network neighborhood who we can regard as stronger ties. In particular, we see that those dyads who have supported connections and are community co-members are more likely to join the same experimental risk pooling group than any other set of dyads. Second, we see through community measures and distance-2 connections that a weaker form of risk pooling tends to extend beyond this neighborhood. This fact is interesting for the collection of networks. In particular, we can detect strong and weak ties even if we measure only one social network of constant intensity. This is particularly useful since network data can be difficult and time consuming to collect. Finally, when network data is at hand and spillovers are present, community detection may be of use to complement other methods in bounding the effect of spillovers mediated by networks. These might be useful for estimating treatment effects themselves, and in providing sanity checks for other assumptions about how spillovers decay in networks.

6.2 New and Unanswered Questions

6.2.1 Community Detection and Economic Networks

While risk pooling is an exciting application of community detection, it may prove valuable for places where networks are relevant to the provision of goods. Similar algorithms have already been used to understand the limits of occupational mobility (Schmutte, 2014). Among other topics, communities may be relevant to the flow of information in economies. Likewise, bipartite communities may be of interest in buyer-seller networks.

6.2.2 Homophily and Network Formation

New questions arise from the community detection measure. Given that detected communities may bound the scope of risk pooling, it becomes interesting how these communities are composed relative to network neighborhoods. In particular, it is often the case that network formation is guided by *homophily*, or the principal that "birds of a feather flock together." Are communities homophilous to the same degree as network neighborhoods?

References

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

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.

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

J.-M. Baland, C. Guirkinger, and C. Mali. Pretending to Be Poor: Borrowing to Escape Forced Solidarity in Cameroon. *Economic Development and Cultural Change*, 60(1):1–16, 2011. ISSN 00130079. doi: 10.1086/661220. URL http://ideas.repec.org/a/ucp/ecdecc/doi10.1086-661220.html%5Cnfiles/147/doi10.1086-661220.html.

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. URL http://economics.mit.edu/files/530.

A. V. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. Changes in Social Network Structure in Response to Exposure to Formal Credit Markets. 2018. doi: 10.2139/ssrn.3245656.

M. Belhaj and F. Deroïan. Risk taking under heterogenous revenue sharing. *Journal of Development Economics*, 98(2):192–202, 2012. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2011.07.003. URL http://dx.doi.org/10.1016/j.jdeveco.2011.07.003.

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

F. Bloch, G. Genicot, and D. Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143(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. ISSN 03043878. doi: 10.1016/j.jdeveco.2016.01.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0304387816000109.

S. Boucher and M. Delpierre. The impact of index-based insurance on informal risk-sharing arrangements. 2014.

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. Public Goods in Networks. *Journal of Economic Theory*, 135(1): 478–494, 2007. ISSN 00220531. doi: 10.1016/j.jet.2006.06.006.

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. ISSN 1556-5068. doi: 10.2139/ssrn.273401.

A. C. Cameron and D. L. Miller. Robust Inference for Dyadic Data. 2014. URL cameron.econ.ucdavis.edu/research[Accessed:10.09.2015].

F. Cecchi, J. Duchoslav, and E. Bulte. Formal insurance and the dynamics of social capital: Experimental evidence from Uganda. *Journal of African Economies*, 25(3):418–438, 2016. ISSN 14643723. doi: 10.1093/jae/ejw002.

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

P.-A. Chiappori, K. Samphantharak, S. Schulhofer-Wohl, and R. M. Townsend. Heterogeneity and risk sharing in village economies. *Quantitative Economics*, 5(1):1–27, 2014. URL http:

//onlinelibrary.wiley.com.ezproxy.library.wisc.edu/doi/10.3982/QE131/abstract.

S. Coate and M. Ravallion. Reciprocity without commitment insurance arrangements. *Journal of Development Economics*, 40:1–24, 1993.

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.

J. de Weerdt. Risk-sharing and endogenous network formation. 2002.

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. ISSN 03043878. doi: 10.1016/j.jdeveco.2015.09.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0304387815001108.

S. Dercon and P. Krishnan. In Sickness and in Health: Risk Sharing within Households in Rural Ethiopia. 108(4), 2001. doi: 10.2139/ssrn.250801.

S. Dercon, J. De Weerdt, T. Bold, and A. Pankhurst. Group-based funeral insurance in Ethiopia and Tanzania. *World Development*, 34(4):685–703, 2006. ISSN 0305750X. doi: 10.1016/j.worlddev.2005.09.009.

P. A. Diamond. American Economic Association 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 Savings on Informal Risk-Sharing : Experimental Evidence from Vulnerable Women in Kenya. *Journal of Human Resources*, 2019.

C. Elbers, J. W. Gunning, and B. Kinsey. Growth and risk: Methodology and micro evidence. *World Bank Economic Review*, 21(1):1–20, 2007. ISSN 02586770. doi: 10.1093/wber/lhl008.

M. Fafchamps. Risk sharing and quasi-credit. *The Journal of International Trade & Economic Development*, 8(3):257–278, 1999. ISSN 0963-8199. doi: 10.1080/09638199900000016. URL http://www.tandfonline.com/doi/abs/10.1080/09638199900000016.

M. Fafchamps. Risk Sharing Between Households. In *Handbook of Social Economics*. 2008.

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. URL http://www.atypon-link.com/AEAP/doi/abs/10.1257/aer.97.2.753://publication/uuid/49D0FF07-BEA7-435F-AA53-A7D77303428B.

M. Fafchamps and S. Lund. Risk-sharing networks in rural Philippines. *Journal of Development Economics*, 2003. ISSN 03043878. doi: 10.1016/S0304-3878(03)00029-4.

E. Fitzsimons, B. Malde, and M. Vera-Hernández. Group Size and the Efficiency of Informal Risk Sharing. *Economic Journal*, 128(612):F575–F608, 2018. ISSN 14680297. doi: 10.1111/ecoj.12565.

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*, 2004. ISSN 1063651X. doi: 10.1103/PhysRevE.69.026113. URL http://arxiv.org/abs/cond-mat/0308217%5Cnhttp://www.arxiv.org/pdf/cond-mat/0308217.pdf.

M. Goldstein. Intrahousehold efficiency and individual insurance in Ghana. *Development Economics Discussion Paper*, pages 1–23, 2004. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127007.

B. S. Graham. *Dyadic regression.* Number August 2019. Elsevier Inc., 2020. ISBN 9780128117712.

doi: 10.1016/b978-0-12-811771-2.00008-0. URL https://doi.org/10.1016/B978-0-12-811771-2.00008-0.

S. H. Heß, D. Jaimovich, and M. Sch. Development Projects and Economic Networks: Lessons From Rural Gambia. (June), 2018.

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. ISSN 00028282. 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. ISSN 0002-8282. doi: 10.1257/aer.102.5.1857.

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 & Proceedings*, 108:422–25, 2018. ISSN 2574-0768. doi: 10.1257/pandp.20181120.

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

D. Karlan, M. Mobius, T. Rosenblat, and A. Szeidl. Trust and Social Collateral. *Quarterly Journal of Economics*, (August), 2009.

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

C. Kinnan. Distinguishing barriers to insurance in Thai villages. *Journal of Human Resources*, 2019.

R. Kranton and Y. Bramoullé. Risk Sharing across Communities. 97(2):70–74, 2011. ISSN 00028282. doi: 10.2307/30034423.

M. P. Leung. Treatment and Spillover Effects Under Network Interference. *Review of Economics and Statistics*, pages 1–42, 2019a. ISSN 0034-6535. doi: 10.1162/rest{_}a{_}00818.

M. P. Leung. Causal Inference Under Approximate Neighborhood Interference. 2019b.

E. Ligon. Risk Sharing and Information in Village Economies. *Review of Economic Studies*, 65(4):847–864, 1998. ISSN 00346527. doi: 10.1111/1467-937X.00071.

E. Ligon and L. Schechter. Measuring Vulnerability. *Economic Journal*, 113(486):95–102, 2011.

E. Ligon, J. Thomas, and T. Worrall. Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies. *Review of Economic Studies*, 69(1):209–244, 2002. ISSN 0034-6527. doi: 10.1111/1467-937X.00204.

M. J. Mccord, J. Roth, and D. Liber. The Landscape of Microinsurance in the World's 100 Poorest Countries. Technical Report April, 2007. URL http://www.microinsurancecentre.org/UploadDocuments/Landscapestudypaper.pdf.

Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444, 2001. ISSN 0360-0572. doi: 10.1146/annurev.soc.27.1.415. URL https://www.jstor.org/stable/2678628?pq-origsite=summon&seq=1#metadata_info_tab_contents.

R. Murgai, P. Winters, E. Sadoulet, and A. d. Janvry. Localized and Incomplete Mutual Insurance. 2000.

M. E. J. Newman. Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1):25–31, 2011. ISSN 1745-2473. doi: 10.1038/nphys2162. URL http://dx.doi.org/10.1038/nphys2162.

P. Pons and M. Latapy. Computing communities in large networks using random walks. *Journal* *of Graph Algorithms Applications*, 10(2):284–293, 2004.

I. M. Schmutte. Free to Move? A Network Analytic Approach for Learning the Limits to Job Mobility. *Labour Economics*, 29:49–61, 2014. ISSN 09275371. doi: 10.1016/j.labeco.2014.05.003. URL http://dx.doi.org/10.1016/j.labeco.2014.05.003.

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. ISSN 15372707. doi: 10.1080/07350015.2017.1409630. URL https://doi.org/10.1080/07350015.2017.1409630.

R. M. Townsend. Risk and Insurance In Village India. *Econometrica*, 62(May):539–591, 1994. ISSN 00129682. doi: 10.2307/2951659.

Figures

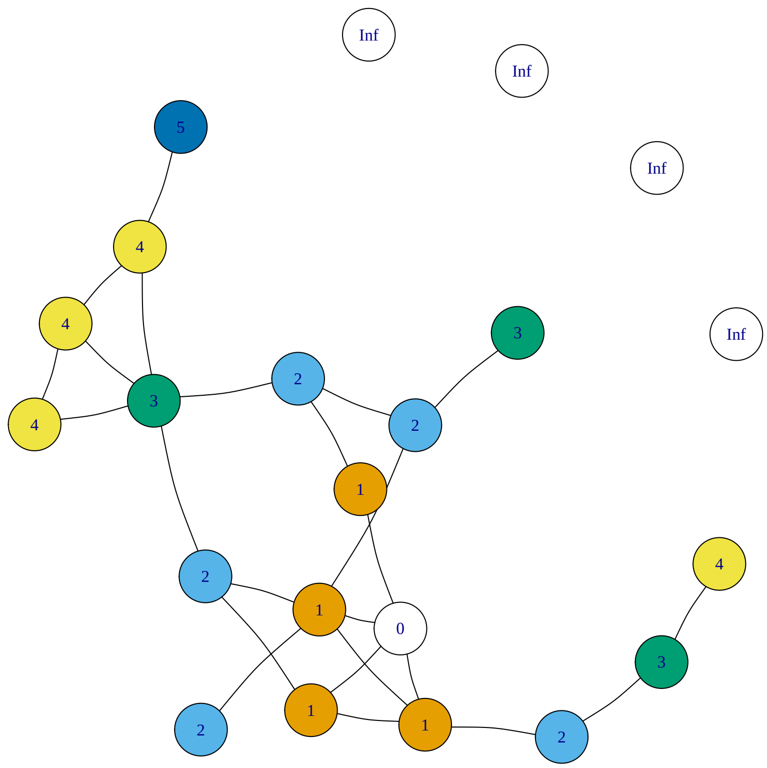

Figure 1: Friends and Family Network with distances from an origin node overlaid. Here 0 is the origin, 1 indicates the set of distance-1 connections, 2 indicates the set of distance-2 connections, and so on.

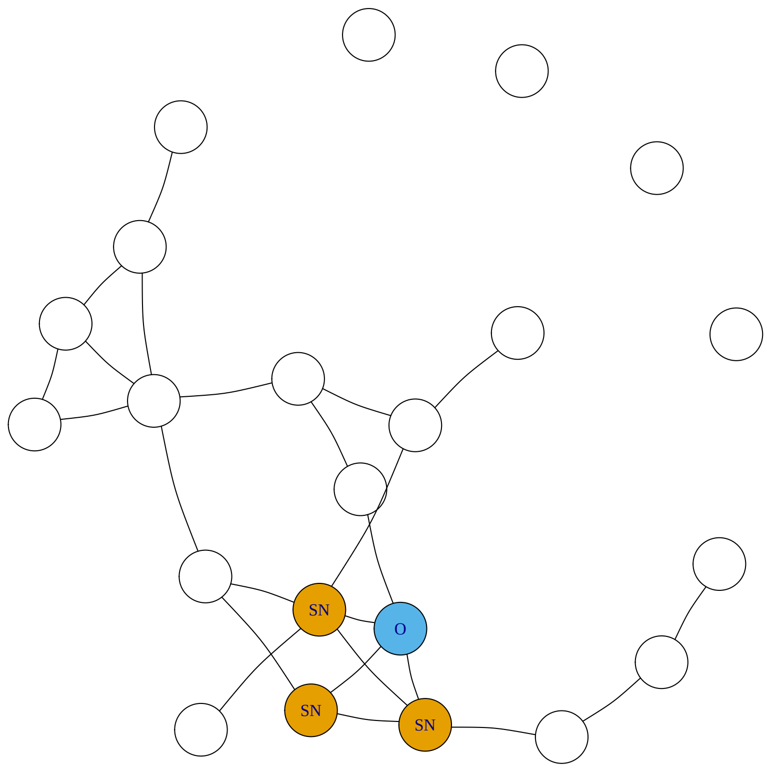

Figure 2: Friends and Family Network with common friends of an origin node overlaid. Here O is the origin and SN indicates the set of supported neighbors.

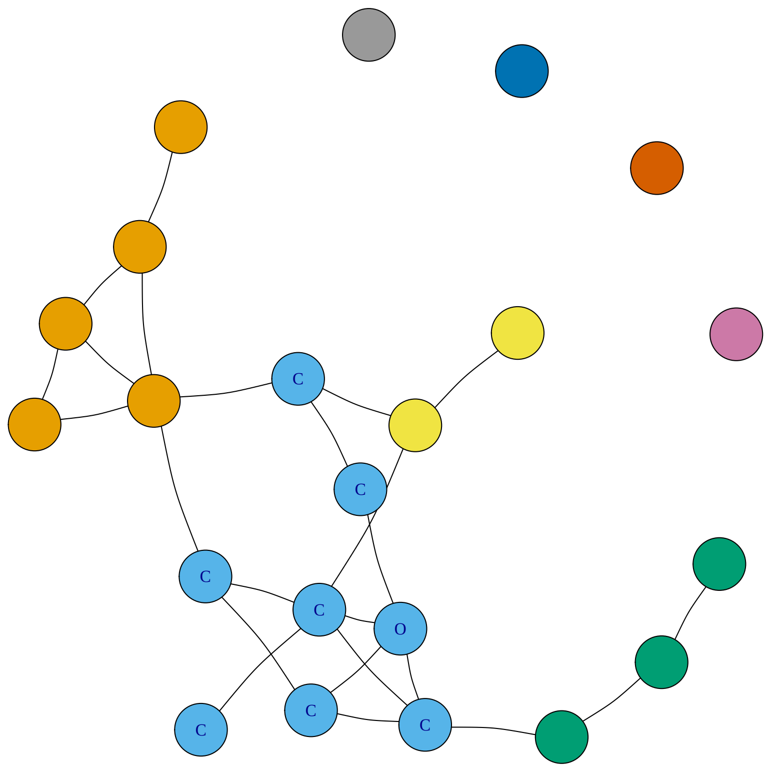

Figure 3: Friends and Family Network with community detection overlaid. Here O is the origin and C indicates those in their detected community.

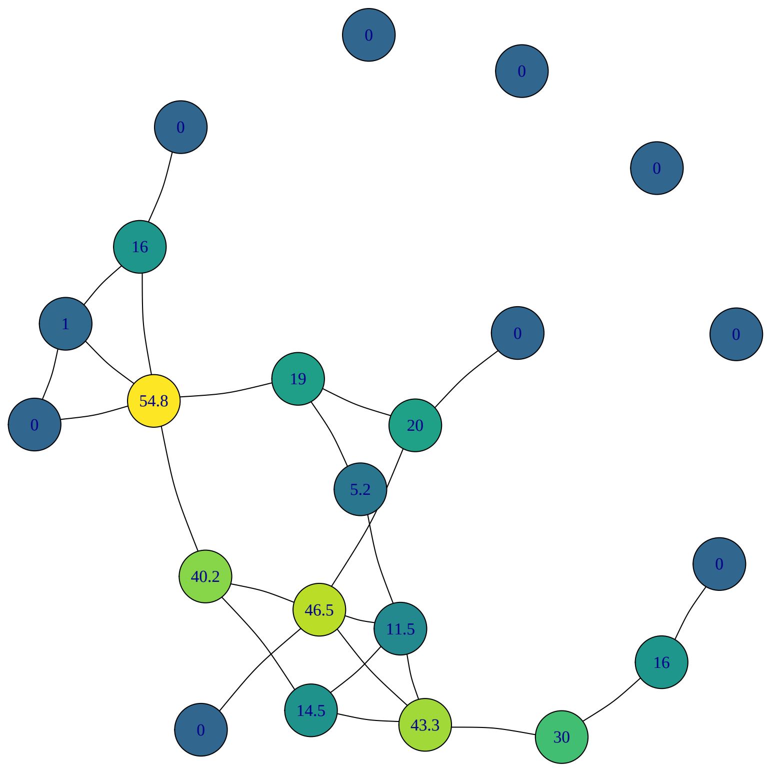

Figure 4: Friends and Family Network with betweenness centrality overlaid. Blue is lowest and yellow is highest, with node betweenness printed on each node.

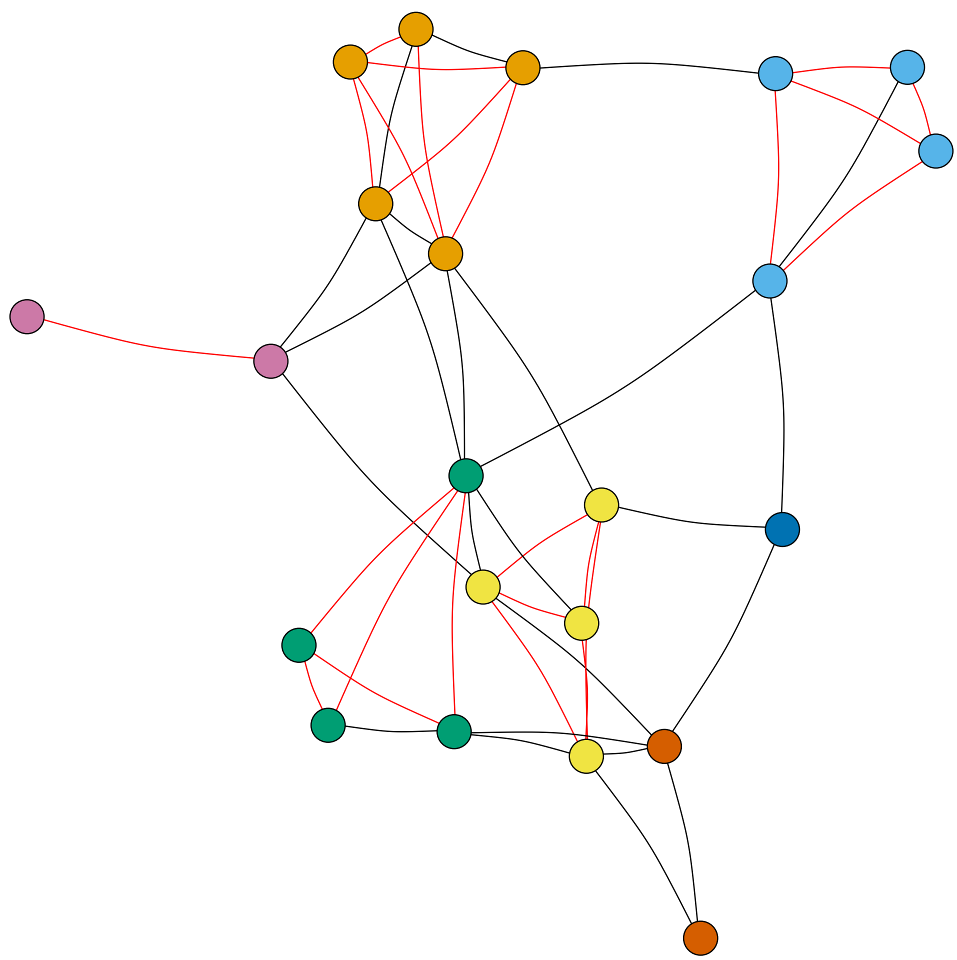

Figure 5: Friends and family and experimental risk pooling groups plotted as a network. Node color is experimental risk pooling group membership. Black edges are connections within the friends and family network, whereas red edges represent co-membership in the risk pooling group where no friends and family network already exists.

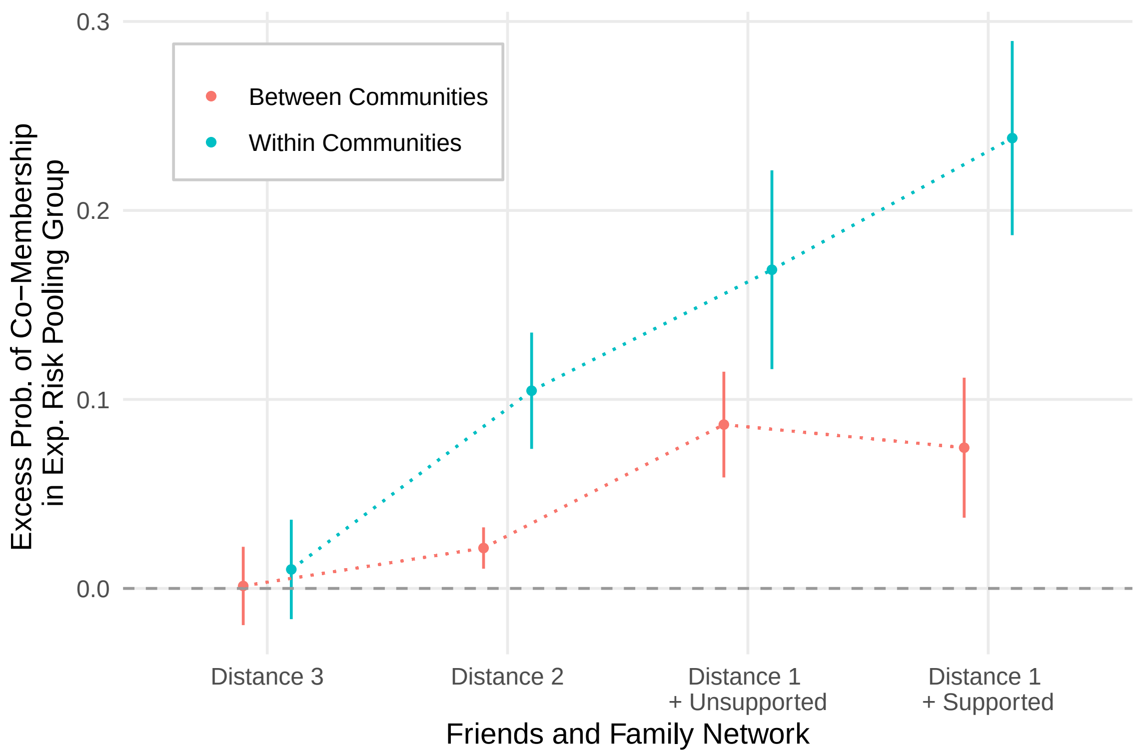

Figure 6: Strong and Weak Ties on the Extensive Margin: Excess probability of dyadic co-membership in an experimental risk pooling group conditional on network relationship.

Tables

Table 1: The Scope of Risk Sharing in the Literature

Paper	Type	Assumption	Result
Dercon and Krishnan (2001)	E	Intra-Household	-
Goldstein (2004)	TE	Intra-Household, Bilateral	
Fafchamps (1999)	T	Bilateral	-
Jack and Suri (2014)	E	Bilateral	-
Fafchamps and Gubert (2007)	E	Bilateral	-
Fafchamps and Lund (2003)	E	Bilateral (Friendship, Kinship)	-
de Weerdt (2002)	E	Bilateral	Network (Common friends)
de Weerdt and Dercon (2006)	TE	Bilateral, Network (2-Shell)	-
Ambrus et al. (2014)	T	Bilateral	Group/Network (Island)
Fitzsimons et al. (2018)	E	Group (Kinship)	Extensive/Intensive Trade-off
Bloch et al. (2008)	T	Bilateral	Network (Flows)
Murgai et al. (2000)	TE	Network (Clusters)	Extensive/Intensive Trade-off
Dercon et al. (2006)	E	Group (Funeral Society)	-
Genicot and Ray (2003)	T	Group	Bounded Group Size
Attanasio et al. (2012a)	TE	Group (Experimental)	Risk Preferences, Friendship, Kinship
Bramoullé and Kranton (2007)	T	Network (Flows - Component)	Bounded Component Size
Kranton and Bramoullé (2011)	T	Network (Flows - Component)	-
Townsend (1994)	TE	Group (Village)	-
Ligon (1998)	TE	Group (Village)	-
Kinnan (2019)	TE	Group (Village)	-
Chiappori et al. (2014)	TE	Group (Village)	-

Type is T if the paper is theoretical, E if empirical, and TE if both. Papers are (imprecisely and qualitatively) ordered by the scope of risk pooling modeled.

Table 2: Incentive Structure for the Gamble Choice Game

		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

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

Table 3: Network Characteristics

Statistic	Friends and Family	Close Friends and Family
Nodes	33.971	
	(11.954)	
Density	0.056	0.026
	(0.044)	(0.022)
Clustering	0.336	0.425
	(0.202)	(0.301)
Closeness	0.547	0.742
	(0.169)	(0.162)
Community size	3.933	2.132
	(2.614)	(1.021)
Modularity	0.429	0.563
	(0.170)	(0.230)

Standard errors in parentheses.

Proximity	Within Community	Between Communities
Supported	$\beta_0 + \beta_1 + \gamma + \delta_0 + \delta_1$	$\beta_0 + \beta_1$
Distance-1	$\beta_1 + \gamma + \delta_1$	β_1
Distance-2	$\beta_2 + \gamma + \delta_2$	β_2
Distance-3	$\beta_3 + \gamma + \delta_3$	β_3
Distance-4+	γ	

Table 4: Excess Probability of Community Co-Membership

Table 5: Dyadic Regressions: Main Specifications, Friends and Family Network

	Co-Membership in Risk Pooling Group						
2-8	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported	0.197***				0.0959***	0.152***	0.0827***
	(10.15)				(4.38)	(9.91)	(4.02)
Friend or Family		0.176***		0.136***	0.104***		0.0759***
		(11.02)		(10.96)	(7.30)		(5.19)
Same Community			0.123***	0.0655***		0.0703***	0.0622***
			(8.91)	(6.17)		(6.98)	(6.07)
Constant	0.0906***	0.0880***	0.0861***	0.0815***	0.0879***	0.0828***	0.0817***
	(59.17)	(53.44)	(38.14)	(34.40)	(53.68)	(35.49)	(35.38)
N	88266	88266	88266	88266	88266	88266	88266
Muni FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, all t tests one sided

Table 6: Dyadic Regressions: Longer Walks and Interactions, Friends and Family Network

	Co-Membership in Risk Pooling Group						
2-8	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported FF			0.0882^{***}	0.0855^{***}			0.00971
			(4.37)	(4.51)			(0.45)
Friends or Family	0.193^{***}	0.148^{***}	0.0872^{***}		0.0833^{***}	0.0935^{***}	0.0901^{***}
	(11.99)	(10.85)	(5.33)		(6.15)	(5.82)	(5.55)
Distance-2 FF	0.0388^{***}	0.0183	0.0226^{*}			0.0226^{*}	0.0251^{*}
	(3.27)	(1.46)	(1.78)			(1.81)	(1.98)
Distance-3 FF	0.00661	0.000132	0.00216			0.00371	0.00546
	(0.64)	(0.01)	(0.20)			(0.35)	(0.51)
Same Community		0.0576^{***}	0.0523^{***}	0.0580^{***}	0.0473^{***}	0.0195	0.0185
		(5.07)	(4.85)	(6.20)	(4.64)	(0.55)	(0.52)
Supported $\times$ Same Comm.				0.0937^{***}			0.101^{**}
				(3.85)			(3.09)
FF $\times$ Same Comm.					0.0870^{***}	0.112^{**}	0.0281
					(4.59)	(2.63)	(0.64)
Distance-2 $\times$ Same Comm.						0.0222	0.0226
						(0.57)	(0.58)
Distance-3 $\times$ Same Comm.						-0.00894	-0.00901
						(-0.25)	(-0.26)
N	88266	88266	88266	88266	88266	88266	88266
Muni FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, all t tests one sided

Table 7: Dyadic Regressions: Main Specifications, Close Friends and Family Network

	Co-Membership in Risk Pooling Group						
2-8	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Close FF Supported	0.240***				0.0967**	0.162***	0.0914**
	(8.67)				(2.87)	(6.24)	(2.80)
Close Friend of Family		0.215***		0.154***	0.146***		0.0903***
		(9.90)		(8.11)	(7.51)		(4.55)
Same Cl. Community			0.156***	0.0716***		0.0907***	0.0698***
			(8.51)	(4.23)		(5.61)	(4.11)
N	88266	88266	88266	88266	88266	88266	88266
Muni FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

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

Table 8: Dyadic Regressions: Longer Walks and Interaction Effects, Close Friends and Family Network

	Co-Membership in Risk Pooling Group						
2-8	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported Cl. FF			0.0810	0.0830**			0.0303
			(1.91)	(2.89)			(0.89)
Close Friends or Family	0.191***	0.158***	0.102***		0.0459	0.0564**	0.0401
	(8.38)	(8.71)	(4.02)		(1.57)	(3.09)	(1.63)
Distance-2 Cl. FF	0.0670***	0.0439***	0.0223			0.0115	0.00959
	(5.49)	(3.47)	(1.24)			(0.58)	(0.47)
Distance-3 Cl. FF	-0.0233	-0.0355	-0.0319			-0.0212	-0.0204
	(-1.39)	(-1.74)	(-1.62)			(-0.98)	(-0.93)
Same Cl. Community		0.0647*	0.0756*	0.0574**	0.0571**	0.00765	0.0260
		(2.25)	(2.40)	(3.38)	(3.39)	(0.26)	(0.89)
Supported Cl. FF $\times$ Same Cl. Comm.				0.117***			0.121***
				(5.45)			(4.85)
Cl. FF $\times$ Same Cl. Comm.					0.132***	0.128***	
					(3.58)	(4.83)	
Disance-2 Cl. FF $\times$ Same Cl. Comm.						0.0777*	0.0637*
						(2.51)	(2.27)
Disance-3 Cl. FF $\times$ Same Cl. Comm.						-0.00612	-0.0135
						(-0.18)	(-0.39)
N	88266	88266	88266	88266	88266	88266	88266
Muni FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

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

Table 9: Dyadic Regressions: Main Effects, Controls

	Co-Membership in Risk Pooling Group						
2-8	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported FF	0.201***				0.102***	0.167***	0.0947***
	(10.21)				(4.79)	(9.92)	(4.58)
Friends and Family		0.176***		0.146***	0.104***		0.0805***
		(10.78)		(10.49)	(7.17)		(5.52)
Same community			0.109***	0.0592***		0.0655***	0.0569***
			(8.21)	(5.90)		(6.71)	(5.80)
N	88266	88266	88266	88266	88266	88266	88266
Muni FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 10: Dyadic Regressions: Longer Walks and Interaction Effects, Controls

	Co-Membership in Risk Pooling Group						
2-8	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Supported FF			0.0761***	0.108***			0.00858
			(3.78)	(5.66)			(0.40)
Friends and Family	0.172***	0.151***	0.0971***		0.0963***	0.0996***	0.0950***
	(11.20)	(11.47)	(6.61)		(7.08)	(7.36)	(6.64)
Distance-2 FF	0.0513***	0.0399***	0.0316***			0.0290***	0.0289***
	(8.26)	(6.61)	(5.51)			(5.00)	(5.07)
Distance-3 FF	0.00502	0.000110	0.00440			0.00881	0.00958
	(0.48)	(0.01)	(0.42)			(0.81)	(0.88)
Same Community		0.0438***	0.0442***	0.0547***	0.0423***	0.00887	0.0466*
		(4.10)	(4.12)	(6.40)	(4.71)	(0.42)	(2.32)
Supported FF $\times$ Same Comm.				0.0814**			0.0685*
				(3.33)			(2.12)
FF $\times$ Same Comm.					0.0802***	0.0847***	0.0276
					(4.34)	(4.71)	(1.23)
Distance-2 $\times$ Same Comm.						0.0640***	0.0319*
						(4.01)	(2.35)
Distance-3 $\times$ Same Comm.						-0.0335	-0.0460*
						(-1.66)	(-2.34)
N	88266	88266	88266	88266	88266	88266	88266
Muni FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

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

Table 11: Defaults by Group, Friends and Family Network

	Proportion of Defaults in Risk Pooling Group				
2-6	(1)	(2)	(3)	(4)	(5)
Group size	0.00350	0.00279	0.00314	0.00524	-0.000354
	(0.93)	(0.66)	(0.71)	(1.00)	(-0.10)
Supported FF	0.0529				
	(0.71)				
Group size × Supported FF	-0.0134				
	(-0.71)				
Friends and family		0.0364			
		(0.48)			
Group size × Friends and Family		-0.00769			
		(-0.41)			
Distance-2			0.0130		
			(0.25)		
Group size × Distance-2			-0.00513		
			(-0.42)		
Distance-3				0.0377	
				(0.71)	
Group size × Distance-3				-0.00763	
				(-0.65)	
Community					-0.0115
					(-0.24)
Group size × Community					0.00409
					(0.37)
Constant	0.242	0.243	0.250	0.225	0.243
	(0.75)	(0.75)	(0.78)	(0.70)	(0.76)
N	526	526	526	526	526
Muni FE	Yes	Yes	Yes	Yes	Yes

Outcome is proportion of defaults and all network variables are computed as densities.

t statistics in parentheses.

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

Table 12: Defaults by Group, Close Friends and Family Network

	Proportion of Defaults in Risk Pooling Group				
2-6	(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 Cl. FF	-0.129*				
	(-2.63)				
Group size × Supported Cl. FF	0.0293				
	(1.90)				
Close Friends and Family		-0.155**			
		(-3.25)			
Group size × Cl. FF		0.0390*			
		(2.61)			
Distance-2 Cl.			-0.118*		
			(-2.46)		
Group size × Distance-2 Cl.			0.0323*		
			(2.61)		
Distance-3 Cl.				-0.114*	
				(-2.40)	
Group size × Distance-3 Cl.				0.0302*	
				(2.42)	
Cl Comm.					-0.120*
					(-2.39)
Group size × Cl Comm.					0.0282*
					(2.15)
Constant	0.266	0.265	0.263	0.265	0.282
	(0.84)	(0.84)	(0.82)	(0.84)	(0.89)
N	526	526	526	526	526
Muni FE	Yes	Yes	Yes	Yes	Yes

Outcome is proportion of defaults and all network variables are computed as densities.

t statistics in parentheses.

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

Community Detection

Modularity

To compute modularity let k_i and k_j be the degrees of nodes i and j, respectively. Let m be the number of edges in the graph. The expected number of edges between i and j from this rewiring is equal to $k_i k_j/(2m-1) \approx k_i k_j/2m$ ($2m$ since each link has two "stubs", so to speak). I can then compare this expected number of links between i and j to the actual connections: letting A_{ij} be the ijth entry of the adjacency matrix (defined $A_{ij} = \mathbf{1}(ij \in g)$), I take the difference these two numbers:

$$A_{ij} - \frac{k_i k_j}{2m}$$

I can interpret this as connections over expected conditional on node pair degrees. Then, letting C_i be the community membership of node i, and connections over expectation are weighted by the function C_{ij}, where $C_{ij} = \mathbf{1}(C_i, C_j)$. Finally I aggregate to the graph level and normalize by twice the number of links present:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) C_{ij}$$

This serves as an easily computable and straightforward measure of the internal quality of communities (Newman, 2011).

Hacking Community Detection for Risk Sharing

Consumption Smoothing

Another approach to finding risk sharing communities combines panel data on consumption and income with . Essentially, we build a dendrogram using the edge betweenness community detection method but instead of using the off the shelf tuning statistic (i.e., modularity) to cut the dendrogram, we use a risk sharing statistic. One way to this would be on actual risk sharing. With income and consumption at the individual or household level, estimate a risk pooling equation at all possible cuts of the dendrogram.

$$c_{it} = \alpha y_{it} + \gamma_{gt} + \epsilon_{it} \tag{16}$$

where c_{ij} is consumption, y_{it} is income (or income shocks), γ_{gt} are community time fixed effects, and ϵ_{it} is the error. In principle, we then choose the cut where $\hat{\alpha}$ ceases to fall.[25] To the degree the community assignments correspond between this algorithm and the off the shelf method, this should increase our confidence in using off the shelf network methods. This however, still leaves many questions unanswered. Is the best approach to choose a minimum tolerance in the change in $\hat{\alpha}$, or would a penalty on the number of communities serve our purposes better?

[25] Why not the minimum value of $\hat{\alpha}$? Consider the case where you split a community with perfect risk sharing in two communities. The smaller communities will also display perfect risk sharing!

Modularity and Transfer Data

Another possible approach using real risk sharing data is valuable when one can actually see networks and transfers separately. This might build a network where transfers have actually taken place and compute modularity on this auxiliary network. Call τ the transfer network, where $ij \in \tau$ if either agent i or j have made a transfer to the other. Defining $T_{ij} = \mathbf{1}(ij \in \tau)$, then I can re-write modularity as follows

$$Q(\tau) = \frac{1}{2m} \sum_{ij} \left(T_{ij} - \frac{k_i(\tau) k_j(\tau)}{2m} \right) C_{ij}.$$

This may also be useful to handle larger scale networks like call data networks with transfers. Notably, mobile transfers are much more sparse than voice and SMS calls.